Two problems are considered in this thesis. The first is concerned with correlation model risk and the second with non-Gaussian factor modelling of asset returns.

One of the fundamental problems in the application of mathematical finance results in a real world setting, is the dependence of mathematical models on parameters (correlations) that are hard to observe in markets. The common term for this problem is model risk. The first part of this thesis aims to provide some building blocks in the estimation of the sensitivities of mathematical objects (prices) to correlation inputs. In high dimensions, computational complexities increase faster than exponentially. A typical approach to deal with this problem is to introduce a principal component approach for dimension reduction. We consider the price of portfolios of options and approximations obtained by modifying the eigenvalues of the covariance matrix, then proceed to find analytical upper bounds on the magnitude of the difference between the price and the approximation, under different assumptions. Monte Carlo simulations are then used to plot the difference between the price and the approximation.

In the second part of this thesis, the assumptions and estimation methods of four different factor models with time varying parameters are discussed. These models are based on Sharpe’s single index model. The first model assumes that residuals follow a Gaussian white noise process, while the other three approaches combine the structure of a single factor model with time-varying parameters, with dynamic volatility (GARCH) assumptions on the model components. The four approaches then are used to estimate the time-varying alphas and betas of three different hedge fund strategies. Results are compared.
To my parents
Acknowledgements

First and foremost, I would like to thank my advisor Prof. Luis A. Seco for his guidance, encouragement, and financial support during these years of graduate studies, but most importantly, for believing in me.

I would also like to thank the graduate students, faculty, and staff members of the Department of Mathematics of University of Toronto.

Thanks to Prof. Matt Davison, Prof. Adrian Nachman, Prof. Robert L. Jerrard, Prof. Adam Stinchcombe, Prof. Catherine Sulem and Prof. Matilde Marcolli for their insightful comments and suggestions.

Thanks to my fellow Risklab students, especially to Chenyu Guo.

Thanks to Jonathan Mostovoy for proofreading my work.

I would also like to thank an amazing group of people that made this experience much more enjoyable: Ivan Salgado, Jemima Merisca, Elizabeth Hill, Marcos Escobar, Selim Tawfik, Ling-Sang Tse, Francisco Guevara, my brother Nicolas, my uncle Julio, and Michael Glas.

Finally, I would like to thank my parents Marivi and Nicolas, for their unconditional love and support.
## Contents

1 Introduction  
   1.1 Literature review  
   1.2 Contributions  
   1.3 Future research  

2 Estimation of correlation model risk  
   2.1 Introduction  
   2.2 Stochastic calculus  
   2.3 Black-Scholes theory  
      2.3.1 The Black-Scholes model  
      2.3.2 The Multivariate Black-Scholes model  
   2.4 Upper bounds  
      2.4.1 Non-degenerate Gaussian densities. Mean value theorem.  
      2.4.2 $L^1$ distance inequalities  
      2.4.3 Uncorrelated underlying assets  
      2.4.4 Correlated underlying assets (Two dimensional case)  
   2.5 Conclusion  

3 Gaussian versus non Gaussian factor models  
   3.1 Introduction  
   3.2 Linear regression and estimation methods  
      3.2.1 Least squares estimation  
      3.2.2 Maximum likelihood estimation  
   3.3 Time series and stationary processes  
   3.4 Financial series and GARCH models  
   3.5 Model approaches  
      3.5.1 Model 1 (Gaussian white noise residuals)  
      3.5.2 Model 2 (GARCH residuals)  
      3.5.3 Model 3 (Unobserved GARCH)  
      3.5.4 Model 4 (Weak GARCH residuals)  
   3.6 Applications to hedge fund return modelling  
   3.7 Conclusion  

A Graphs
List of Tables

3.1 Average alpha per period ......................................................... 86
3.2 Average beta per period ......................................................... 87
List of Figures

2.1 Magnitude of the difference between the price of a portfolio of 2 put options and the approxima-
tion obtained after modifying the smallest eigenvalue without making it zero, and the proposed
upper bound. .............................................. 28
2.2 Magnitude of the difference between the price of a digital option and the approximation obtained
by making the variance zero. .............................................. 35
2.3 Magnitude of the difference between the price of a call option and the approximation obtained by
making the variance zero. .............................................. 36
2.4 Magnitude of the difference between the price of a portfolio of 2 digital options and the approxi-
mation obtained by making the smallest eigenvalue zero. .............................................. 52
2.5 Magnitude of the difference between the price of a portfolio of 2 call options and the approxima-
tion obtained by making the smallest eigenvalue zero. .............................................. 52

3.1 Time-varying beta of the equity hedge strategy for models 1, 2 and 4. ......................... 84
3.2 Time-varying alpha of the equity hedge strategy for models 1, 2 and 4. ......................... 85
3.3 Time-varying beta from model 4 for each strategy. ............................................... 87
3.4 Time-varying alpha from model 4 for each strategy. ............................................... 87

A.1 Time-varying beta of the event driven strategy for models 1, 2 and 4. ......................... 89
A.2 Time-varying alpha of the event driven strategy for models 1, 2 and 4. ......................... 90
A.3 Time-varying beta of the merger arbitrage strategy for models 1, 2 and 4. ......................... 90
A.4 Time-varying alpha of the merger arbitrage strategy for models 1, 2 and 4. ......................... 91
A.5 Time-varying beta of each strategy for model 3. ................................................. 91
A.6 Time-varying alpha of each strategy for model 3. ................................................. 92
Chapter 1

Introduction

1.1 Literature review

Finance is one of the fastest developing areas in modern banking and the corporate world. This growth brings along the emergence of sophisticated financial products and the resulting new mathematical models and methods. Several assets are traded daily on financial markets around the world: stocks and bonds are some of the most common ones, however, more complex instruments with payoffs dependent on the price of other assets are now commonly traded too: forward contracts and options. In particular, European options give the holder of the contract the right, but not the obligation, to exercise the option at some future time specified in the contract. Different theories have been proposed through the years to model the dynamics of asset prices, particularly relevant is Samuelson, P. (1965) work. Black, F. and Scholes, M. (1973) derived an equation known as the Black-Scholes equation to price derivatives on a single asset within the Black-Scholes model framework, an adjusted Samuelson market environment; that same year Merton, R. (1973) showed that a Black-Scholes type model can be derived from weaker assumptions than in its original formulation. Cox, J., Ingersoll J. and Ross, S. (1985) extended the Black-Scholes equation to the generalized Black-Scholes equation to price derivatives on multiple assets.

Financial market data often exhibit volatility clustering, where time series show periods of high volatility and periods of low volatility. In fact, in economic and financial data, time varying volatility is more common than constant volatility, therefore accurate modeling of time varying volatility is of great importance. A groundbreaking paper was proposed by Engle, R (1982) to model the heteroscedastic (time varying) conditional variance of a time series as a linear function of past squared observations. This formulation is called Autoregressive Conditional Heteroscedastic (ARCH) model. Bollerslev, T (1986) proposed the generalized ARCH, or GARCH model, which specified the conditional variance to be a function of lagged squared observations and past conditional variance. Several extensions of the GARCH model have been proposed in the literature. The (ARCH-M) model is one of the most popular ones, in this model the conditional variance of the asset is included in the equation of the conditional mean, see Engle, R., Lilien, D. and Robins, R. (1987). Empirical evidence of stock returns shows that a stock price decrease tends the increase subsequent volatility by more than a stock price increase, this phenomenon is called leverage effect. Different models have been suggested to allow for asymmetric effects of positive and negative shocks to volatility, some of the more popular ones are the GJR-GARCH model introduced by Glosten, L., Jagannathan, R. and Runkle, E. (1993) and the Exponential GARCH (EGARCH) introduced by Nelson, D. (1991). Drost, F. and Nijman, T. (1993) showed that temporal aggregation of GARCH processes is in general not GARCH but weak GARCH, a class of processes that they defined in their paper. Nijman, T. and Sentana, E. (1996)
showed that contemporaneous aggregation of independent univariate GARCH processes yields a weak GARCH process. Franq, C. and Zakoian, J. (2000) considered weak GARCH (or weak ARMA-GARCH) representations characterized by an ARMA structure of the squared error terms and proposed a two stage least squares method to estimate the model parameters.

A large part of modern investment theory is based on the idea that asset returns are driven by a small set of basic variables or factors. Factor models are widely used in risk management and portfolio optimization to predict returns, generate estimates of abnormal return, and estimate the variability and co-variability of returns. The three main types of multifactor models for asset returns are: macroeconomic factor models, fundamental factor models, and statistical factor models. The most widely used macroeconomic factor model is Sharpe’s single index model, which was proposed by Sharpe, W. (1964). A multifactor extension was later proposed by Chen, N.F., Roll, R, and Ross, S.A. (1986). The two main fundamental factor models were introduced by Fama E. and French, K.R. (Fama-French model) in 1992 and Rosenberg, B. (Barra model) in 1975 respectively. Roll, R. and Ross, S.A. (1980) provided an early empirical application of statistical factor models to equity return data modelling, later on Connor, G. and Korajczyk, R. (1986), Stock, J.H. and Watson M.W. (2002) and Bai, J. (2003) extended the theory of statistical factor models.

Hedge funds are one of the largest asset classes in the world. They seek to generate positive alpha for investors through volatility and risk reduction. Hedge fund returns are labelled as absolute return investment vehicles because of their ability to generate stable returns with low systematic risk, which has lead to a significant increase of their popularity over the last 15 years. In 1992, Sharpe proposed asset class factor modeling for mutual funds and showed that by using a limited number of asset classes, it was possible to explain the sources of performance for US mutual funds, see Sharpe, W. (1992). Sharpe’s model is, however, less effective for hedge funds, which employ dynamic trading strategies. Fung, W. and Hsieh, D. (1997) employed a multi-factor approach, using three equity classes, two bond classes, commodities and currencies to model hedge fund returns. Criticism that non-linearities were not being captured through simple linear regression models provided motivation for assessing non-linear factors. Non-linear factors were implemented by Agarwal, V . and Naik, N. (2004) and Fung, W. and Hsieh, D. (2004) in their multifactor models. Roncalli, T. and Weisang, G. (2008) pointed out that dynamic trading of assets would in itself result in non-linear return profiles. Hence, instead of specifying option based factors, they highlighted the importance of capturing time-varying beta, because hedge fund strategies are dynamic, therefore time invariant factor loadings are unrealistic and simplistic. Research in time varying beta lead to the next step in hedge fund return modeling, which is replication of hedge fund returns via asset based style factor modeling. Hasan hodzic, J. and Lo, A. (2007), and Darolles, S. and Mero, G. (2007), estimated time varying betas through rolling window regression, using 24-months and 36-months respectively. However, the length of the window affects significantly the estimation results, as longer windows tend to provide more statistical accuracy but are less able to capture recent exposures. The Kalman filter has been suggested to overcome issues with the rolling window approach. Takahashi, A. and Yamamoto, K. (2008) applied both methods: rolling window and Kalman filters to estimate time varying exposures. They noted that on average the Kalman filter captures exposure earlier than rolling window. Similar results were obtained by Wei, W. (2010), who also compared factor based modelling via rolling window against Kalman filters and showed that certain hedge fund strategies are more susceptible to be cloned. Wei provided an excellent review of the evolution of hedge fund return modelling.
1.2 Contributions

Two problems are considered in this thesis. First, given a portfolio of single asset options with correlated underlying assets we study how changes in the eigenvalues of the covariance matrix affect the price of the portfolio. Second, four different one factor rolling window regression approaches are discussed and used to model the returns of three different hedge fund strategies, the time varying alphas and betas of the returns are estimated and results are compared.

One of the fundamental problems in the application of mathematical finance results to a real world setting is the dependence of mathematical models on parameters that are hard to observe in markets. The common term for this problem is model risk. Chapter 2 aims to provide some building blocks in the estimation of the sensitivities of mathematical objects (prices) to correlation inputs. We study the dependence of prices on correlation assumptions or correlation observation errors. In high dimensions, computational complexities increase faster than exponentially, a typical approach to deal with this problem is to introduce a principal component approach for dimension reduction. This approach aims to replace the original covariance (or correlation) matrix by another one where the smaller eigenvalues have been set to 0, thereby reducing the effective dimension of the problem and achieving practical computational efficiency. In particular, we consider the price of portfolios of single asset options, obtained under the assumptions of the Black-Scholes multidimensional model (constant and deterministic parameters). The eigenvalues of the covariance matrix are modified and analytical upper bounds of the magnitude of the difference between the price and the approximation are obtained, for each of the following cases: when the option payoffs are bounded and eigenvalues are modified without making them zero; when the assets are uncorrelated and some eigenvalues are made zero, and the most interesting case, when the portfolio consists of two call or two digital options (both cases where considered) with correlated underlying assets and one of the eigenvalues is set to 0. Monte Carlo simulations are used to plot the magnitude of the difference between the price and the approximation.

In chapter 3, the assumptions and estimation methods of four different single factor models with time varying alpha and beta are discussed. All the considered models use a rolling window for the estimation of time varying parameters. Model 1 makes the traditional assumption that residuals follow a Gaussian white noise process. The other three models combine the structure of a 1-factor model with time varying parameters, with dynamic volatility assumptions on the model components. Model 2 considers a 1-factor model with GARCH(1,1) residuals, model 3 assumes that the centered regressor (market return) is GARCH(1,1) and the residuals are Gaussian white noise, and model 4 assumes that the residuals follow a weak GARCH(2,2) process. To the best of our knowledge the last approach has not been used before to estimate time varying parameters of factor models. Time varying alphas and betas of three different hedge fund strategies are then estimated using each model. Finally, results are compared.

1.3 Future research

Dimension reduction methods play a key role in multi-asset derivatives pricing, therefore it is important to clearly understand how accurate approximations produced by these methods are. In this thesis, analytical upper bounds of the magnitude of the difference between the price and an approximation obtained by making the smallest eigenvalue of the covariance matrix zero, were obtained for portfolios of two options. However, the more general problem, when a portfolio consists of an arbitrary number $n > 2$ of single asset options and $k < n$ eigenvalues are set to 0 still remains open.

In the second part of this thesis we considered four different approaches to estimate the time varying parame-
ters of 1-factor models. It would be interesting however, to extend our analysis to multi-factor models with time varying parameters.
Chapter 2

Estimation of correlation model risk

2.1 Introduction

One of the fundamental problems in the application of mathematical finance results in a real world setting, is the dependence of mathematical models on parameters that are hard to observe in markets. The common term for this problem is model risk, and its study has become popular in recent years. Perhaps the best example of a disregard for model risk was the cause of the financial crisis of 2008: mathematical models that explained CDO prices, while correct from a mathematical perspective, made assumptions on correlations and were used in credit rating and portfolio valuation applications without giving serious thought to the dependence of conclusions on correlation estimations. Correlation matrices are notoriously hard to observe in practice, as historical estimators require extremely long time series when the dimension is high, and correlation numbers implied from market data offer a very reduced field of vision.

This chapter aims to provide some building blocks in the estimation of the sensitivities of prices to correlation inputs. Our motivation is, on the one hand, we are interested, from a risk management perspective, in understanding the dependence of prices on correlation assumptions or correlation observation errors, but also with an eye to assisting in the creation of new mathematical models in high dimensions. Indeed, in high dimensions, computational complexities increase faster than exponentially, and a typical approach is to introduce a principal component approach for dimension reduction. This approach aims to replace the original covariance (or correlation) matrix by another one where the smallest eigenvalues have been set to 0, thereby reducing the effective dimension of the problem and achieving practical computational efficiency. While the process of dimension reduction is common, we want to provide some element of error analysis in the always difficult setting of understanding sensitivities of calculated objects (prices, mainly) to correlation model inputs.

We consider portfolios of single asset options with correlated underlying assets. The multivariate Black-Scholes model framework can be used to model their prices, however as the number of options in the portfolio increases finding the portfolio’s value becomes computationally very expensive and time consuming. In order to solve this problem, dimension reduction methods are commonly used. We approximate the price of portfolios of single asset options by modifying the eigenvalues of the covariance matrix, and obtain analytical upper bounds of the magnitude of the difference between the price and the approximation.

This chapter is organized as follows: section 2.2 provides basic definitions and important results of stochastic processes that will be used in the derivation of the Black-Scholes pricing formula in section 2.3. Upper bounds on the change in price after eigenvalues of the covariance matrix are modified are obtained in section 2.4.
conclusions of this chapter are in section 2.5.

2.2 Stochastic calculus

**Definition 2.2.1.** A probability space \((\Omega, \mathcal{F}, P)\) is said to be a complete probability space if for all \(B \in \mathcal{F}\) with \(P(B) = 0\) and all \(A \subset B\), we have that \(P(A) = 0\).

From now on we assume that all our probability spaces are complete.

**Definition 2.2.2.** Let \((\Omega, \mathcal{F}, P)\) be a probability space. A stochastic process \(X = (X_t)_{t \in [0,T]}\) is a parametrized collection of random variables defined on \((\Omega, \mathcal{F}, P)\) which take values in \(\mathbb{R}^n\).

Note that for each \(t \in [0, T]\) fixed we have a random vector \(\omega \rightarrow X_t(\omega)\).

On the other hand, fixing \(\omega \in \Omega\) we have a function \(t \rightarrow X_t(\omega)\), which is called a path of \(X_t\).

**Definition 2.2.3.** Let \(P\) and \(Q\) be two probability measures defined on the same space \((\Omega, \mathcal{F})\), if for any \(A \in \mathcal{F}\)

\[ P(A) > 0 \Leftrightarrow Q(A) > 0 \]

we say that \(P\) and \(Q\) are equivalent.

**Definition 2.2.4.** Let \((\Omega, \mathcal{F}, P)\) be a probability space. A continuous stochastic process \((W_t)_{t \in [0,T]}\) is said to be Brownian motion if

1. \(W_0 = 0\),
2. \(W_t - W_s\) is independent of \(W_{t'} - W_{s'}\) for \(0 \leq s' < t' \leq s < t \leq T\),
3. \(W_t - W_s \sim N(0, t - s)\) for \(0 \leq s < t \leq T\).

**Definition 2.2.5.** Let \((W_t)_{t \in [0,T]}\) be \(n\)-dimensional Brownian motion. We define \(\mathcal{F}_t = \mathcal{F}_t^{(n)}\) to be the \(\sigma\)-algebra generated by the random variables \((W_{i,s})_{1 \leq i \leq n, 0 \leq s \leq t}\). In other words, \(\mathcal{F}_t\) is the smallest \(\sigma\)-algebra containing all sets of the form \(\{ \omega : W_{t_k}(\omega) \in F_1, \ldots, W_{t_j}(\omega) \in F_j \}\), where \(t_j \leq t\) and \(F_j \subset \mathbb{R}^n\) are Borel sets, \(j \leq k = 1, 2, \ldots\) (We assume that all sets of measure zero are included in \(\mathcal{F}_t\)).

One can think of \(\mathcal{F}_t\) as “the history of \(W_s\) up to time \(t\)”. Note that \(\mathcal{F}_s \subset \mathcal{F}_t\) for \(s < t\) (i.e. \((\mathcal{F}_t)_{t \in [0,T]}\) is increasing) and that \(\mathcal{F}_t \in \mathcal{F}\).

**Definition 2.2.6.** Let \((N_t)_{t \in [0,T]}\) be an increasing family of \(\sigma\)-algebras of subsets of \(\Omega\). A stochastic process \(g_t(\omega) : [0, T] \times \Omega \rightarrow \mathbb{R}^n\) is called \(N_t\)-adapted if for each \(t \in [0, T]\) the function \(\omega \rightarrow g_t(\omega)\)

is \(N_t\)-measurable.
**Definition 2.2.7.** A filtration (on \((\Omega, F)\)) is a family of \(\sigma\)-algebras \(M = (M_t)_{t \in [0, T]}\) with \(M_t \subset F\) for each \(t \in [0, T]\) and such that:

\[M_s \subset M_t \text{ for } 0 \leq s < t.\]

**Remark.** Note that the family of \(\sigma\)-algebras \((F_t)_{t \in [0, T]}\) from definition (2.2.5) is a filtration.

**Definition 2.2.8.** A filtered probability space is a probability space equipped with a filtration.

**Definition 2.2.9.** An \(n\)-dimensional stochastic process \(X = (X_t)_{t \in [0, T]}\) on a filtered probability space \((\Omega, F, \mathbb{P}, M)\) is called a martingale with respect to the filtration \(M = (M_t)_{t \in [0, T]}\) if

1. \(X_t = M_t\) measurable for all \(t \in [0, T]\),
2. \(\mathbb{E}[|X_t|] < \infty\) for all \(t \in [0, T]\),
3. \(\mathbb{E}[X_s | M_t] = X_t\) for all \(0 \leq t < s \leq T\).

**Proposition 1.** A Brownian motion \((W_t)_{t \in [0, T]}\) in \(\mathbb{R}^n\) is a martingale with respect to the family of \(\sigma\)-algebras \((F_t)_{t \in [0, T]}\) generated by \(\{W_s : s \leq t\}\).


**Proposition 2** (Tower Law). Let \((\Omega, F, \mathbb{P})\) be a probability space, \(G_1 \subseteq G_2 \subseteq F\) sub \(\sigma\)-algebras and \(Y\) a random variable defined on such space. If \(\mathbb{E}[|Y|] < \infty\) then

\[\mathbb{E}[\mathbb{E}[Y|G_2]|G_1] = \mathbb{E}[Y|G_1] \quad \mathbb{P}\text{-a.s.}\]


**Definition 2.2.10.** Consider the filtered probability space \((\Omega, F, \mathbb{P}, (F_t)_{t \in [0, T]}))\) and let \(f_t(\omega) : [0, T] \times \Omega \to \mathbb{R}\) be a process such that:

1. \((t, \omega) \to f_t(\omega)\) is \(B \times F\)-measurable, where \(B\) denotes the Borel \(\sigma\)-algebra on \([0, T]\),
2. \(f_t(\omega)\) is \(F_t\)-adapted,
3. \(\mathbb{P}\left(\int_0^T f_t(\omega)^2 dt < \infty\right) = 1\).

The Itô integral is defined as

\[\int_0^T f_t(\omega)dW_t = \lim_{N \to \infty} \sum_{j=0}^{N-1} f_{t_j}(\omega)(W_{t_{j+1}} - W_{t_j})(\omega)\]

where \(t_j = j \cdot \frac{T}{N}\), \((W_t)_{t \in [0, T]}\) is a one dimensional Brownian motion and the limit is in probability.

**Definition 2.2.11.** Let \(W = (W^1, \cdots, W^n)\) be an \(n\)-dimensional Brownian motion defined on the filtered probability space \((\Omega, F, \mathbb{P}, (F_t)_{t \in [0, T]}))\) and \(\sigma_t = [\sigma^{ij}_t]\) a matrix such that for all \(i = 1, \cdots, m\) and \(j = 1, \cdots, n\), \(\sigma^{ij}_t(\omega)\) is a process such that:

1. \((t, \omega) \to \sigma^{ij}_t(\omega)\) is \(B \times F\)-measurable, where \(B\) denotes the Borel \(\sigma\)-algebra on \([0, T]\),
2. \(\sigma^{ij}_t(\omega)\) is \(F_t\)-adapted,
CHAPTER 2. ESTIMATION OF CORRELATION MODEL RISK

3. \( \mathbb{P} \left( \int_0^T (\sigma_{ij}(\omega))^2 dt < \infty \right) = 1. \)

\( \int_0^T \sigma_i(\omega) dW_t(\omega) \) is a column vector whose \( i \)-th component is the following sum of 1-dimensional Itô integrals:

\[
\sum_{j=1}^n \int_0^T \sigma_{ij}(\omega) dW^j_t(\omega).
\]

**Definition 2.2.12.** Let \( W \) be 1-dimensional Brownian motion on the filtered probability space \( (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in [0,T]}) \). A 1-dimensional Itô process \( X \) on \( (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in [0,T]}) \) is a process of the form

\[
X_t = X_0 + \int_0^t u_s(\omega) ds + \int_0^t v_s(\omega) dW_s(\omega),
\]

where both \( u_t(\omega) \) and \( v_t(\omega) \) are \( \mathcal{B} \times \mathcal{F} \)-measurable and \( \mathcal{F}_t \)-adapted, with

\[
\mathbb{P} \left( \int_0^T v_t(\omega)^2 dt < \infty \right) = 1,
\]

and

\[
\mathbb{P} \left( \int_0^T |u_t(\omega)| dt < \infty \right) = 1.
\]

**Definition 2.2.13.** Let \( W \) be \( m \)-dimensional Brownian motion on the filtered probability space \( (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in [0,T]}) \). Let be \( \mu = (\mu^1, \cdots, \mu^n) \) be an \( n \)-dimensional process and \( \sigma = [\sigma_{ij}] \) an \( n \times m \) matrix of processes such that for all \( i = 1, \cdots, n \) and \( j = 1, \cdots, m \), we have that \( \mu^i_t(\omega) \) and \( \sigma_{ij}^t(\omega) \) are \( \mathcal{B} \times \mathcal{F} \)-measurable and \( \mathcal{F}_t \)-adapted with

\[
\mathbb{P} \left( \int_0^T \sigma_{ij}^t(\omega)^2 dt < \infty \right) = 1,
\]

and

\[
\mathbb{P} \left( \int_0^T |\mu^i_t(\omega)| dt < \infty \right) = 1.
\]

An \( n \)-dimensional Itô process \( X \) on \( (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in [0,T]}) \) is a process of the form

\[
X_t = X_0 + \int_0^t \mu_s(\omega) ds + \int_0^t \sigma_s(\omega) dW_s(\omega).
\]

An Itô process can also be written in the following (differential) form

\[
dX_t = \mu dt + \sigma dW_t.
\]

**Theorem 3 (\( n \)-dimensional Martingale Representation Theorem).** Let \( W \) be \( n \)-dimensional Brownian motion on the probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) and \( \mathcal{F}_t \) be the filtration generated by \( W \). Suppose that \( M \) is an \( n \)-dimensional \( \mathbb{P} \)-martingale process, \( M = (M^1, \cdots, M^n) \) which has volatility matrix \( \sigma \), in that \( dM^i_t = \sum_{j=1}^n \sigma^i_j dW^j_t \), and the matrix satisfies the additional condition that (with probability one) it is always non-singular. Then if \( N \) is any \( 1 \)-dimensional \( \mathbb{P} \)-martingale, then there exists an \( n \)-dimensional \( \mathcal{F}_t \)-adapted process \( \phi = (\phi^1, \cdots, \phi^n) \) such that
the martingale $N$ can be written as

$$N_t = N_0 + \sum_{j=1}^{n} \int_0^t \phi^j_s dM^j_s.$$  

Proof. See Oksendal, B. (2003), pp. 51-54. \hfill $\square$

**Theorem 4** (Itô formula). Let $X$ be a 1-dimensional Itô process given by

$$dX_t = udt + vdW_t.$$  

Let $f(t,x)$ be in $C^2([0, \infty) \times \mathbb{R})$ and define the process $Z$ by $Z_t = f(t, X_t)$. Then $Z$ is again an Itô process and

$$dZ_t = \frac{\partial f}{\partial t}(t, X_t) dt + \frac{\partial f}{\partial x}(t, X_t)dX_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(t, X_t)(dX_t)^2,$$

where $(dX_t)^2$ is computed according to the following rules: $(dt)^2 = 0$, $dt dW_t = dW_t dt = 0$ and $(dW_t)^2 = dt$.

Proof. See Oksendal, B. (2003), pp. 46-48. \hfill $\square$

**Definition 2.2.14.** Let $\overrightarrow{W}$ be a vector of $d$ independent Brownian motion and $\delta = [\delta_{ij}]$ be a deterministic and constant matrix with unit length rows i.e. $||\delta_{ii}|| = 1$ for $i = 1, \cdots, n$, then the vector process $W = \delta \overrightarrow{W}$ is a vector of correlated Brownian motion with correlation matrix $\rho = \delta \delta'$.

**Theorem 5** (General Itô formula). Let $X$ be an $n$-dimensional Itô process given by

$$dX_t = \mu dt + \sigma dW_t$$

where $W = (W^1, \cdots, W^m)$ is a correlated vector of Brownian motion and let $f(t, X_t)$ be a $C^2$ map from $[0, \infty) \times \mathbb{R}^n$ into $\mathbb{R}$, the stochastic differential of the process $Z$, where $Z_t = f(t, X_t)$, is given by

$$dZ_t = \frac{\partial f}{\partial t}(t, X_t) dt + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(t, X_t)dX^i_t + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t)dX^i_t dX^j_t$$

with the formal multiplication table: $(dt)^2 = 0$, $dt dW^i_t = 0$ and $dW^i_t dW^j_t = \rho_{ij} dt$ for all $i, j = 1, \cdots, m$.

In particular, if $m = n$ and $dX$ has the structure

$$dX^i_t = \mu^i dt + \sigma^i dW^i_t,$$

for $i = 1, \cdots, n.$

where $\mu^i$ and $\sigma^i$ are scalar processes, then the stochastic differential of the process $Z$ is given by

$$dZ_t = \left(\frac{\partial f}{\partial t}(t, X_t) + \sum_{i=1}^{n} \mu^i \frac{\partial f}{\partial x_i}(t, X_t) + \frac{1}{2} \sum_{i,j=1}^{n} \sigma^i \sigma^j \rho_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t)\right) dt + \sum_{i=1}^{n} \sigma^i \frac{\partial f}{\partial x_i}(t, X_t)dW^i_t.$$

Proof. See Karatzas, I. and Shevrez, S. (1991), pp. 150-153. \hfill $\square$

**Definition 2.2.15.** Let $W$ be an $m$-dimensional Brownian motion vector, $T > 0$, $\mu(\cdot, \cdot) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$, $\sigma(\cdot, \cdot) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$ be measurable functions. An equation of the form

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t$$
is called a *stochastic differential equation* (SDE).

The following theorem gives conditions for a solution to the initial value problem

\[
\begin{align*}
\frac{dX_t}{dt} &= \mu(t, X_t)dt + \sigma(t, X_t)dW_t, \quad t \in [0, T], \\
X_0 &= x_0.
\end{align*}
\]  

(2.1)

to exist and be unique.

**Theorem 6** (Existence and uniqueness theorem for stochastic differential equations). Suppose that the functions \( \mu \) and \( \sigma \) satisfy that

\[
|\mu(t, x)| + |\sigma(t, x)| \leq C(1 + |x|); \quad x \in \mathbb{R}^n, t \in [0, T]
\]

for some constant \( C > 0 \), (where \( |\sigma|^2 = \sum |\sigma^{ij}|^2 \)) and

\[
|\mu(t, x) - \mu(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y|; \quad x, y \in \mathbb{R}^n, t \in [0, T]
\]

for some constant \( D \), then there exists a unique solution to the SDE in (2.1) that is \( \mathcal{F}_t \)-adapted and such that

\[
E[|X_t|^2] \leq Ke^{Kt}(1 + |x_0|^2); \quad t \in [0, T],
\]

for some constant \( K \).


**Theorem 7** (Girsanov Theorem). Let \( W = (W^1, \ldots, W^n) \) be an \( n \)-dimensional \( \mathbb{P} \)-Brownian motion defined on the filtered probability space \( (\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in [0, T]}) \). Suppose that \( \theta = (\theta^1, \ldots, \theta^n) \) is an \( \mathcal{F}_t \)-adapted vector process which satisfies the Novikov condition: \( E\left[\exp\left(\frac{1}{2} \int_0^T \theta_t^2 ds\right)\right] < \infty \), and we set \( \hat{W}^i_t = W^i_t + \int_0^t \theta^i_s ds \) for \( t \in [0, T] \). Then there exists a measure \( \mathbb{Q} \), equivalent to \( \mathbb{P} \) up to time \( T \), such that \( \hat{W} = (\hat{W}^1, \ldots, \hat{W}^n) \) is an \( n \)-dimensional \( \mathbb{Q} \)-Brownian motion up to time \( T \). The Radon-Nikodym derivative of \( \mathbb{Q} \) by \( \mathbb{P} \) is

\[
\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(-\sum_{i=1}^n \int_0^T \theta^i_s dW^i_s - \frac{1}{2} \int_0^T |\theta_t|^2 dt\right)
\]


### 2.3 Black-Scholes theory

**Definition 2.3.1** (Geometric Brownian motion). A process \( X \) with stochastic differential equation

\[
\frac{dX_t}{dt} = \mu X_t dt + \sigma X_t dW_t
\]

is called geometric Brownian motion.

If \( X \) is a 1-dimensional geometric Brownian motion with SDE

\[
\frac{dX_t}{dt} = u X_t dt + v X_t dW_t,
\]

and initial condition

\[
X_0 = x,
\]
then

\[ X_t = x \exp \left( (u - \frac{1}{2}v^2)t + vW_t \right). \]

This can be verified using Itô formula.

We denote the price process of a risk free asset as \( B \), with dynamics:

\[
\begin{align*}
\frac{dB_t}{B_0} &= rB_t dt, \\
B_0 &= 1.
\end{align*}
\]

where \( r \) is a constant.

### 2.3.1 The Black-Scholes model

**Definition 2.3.2** (Black Scholes model). The Black-Scholes model assumes a market consisting of two assets with dynamics given by

\[
\begin{align*}
\frac{dB_t}{B_t} &= rB_t dt, \\
\frac{dS_t}{S_t} &= \mu S_t dt + \sigma S_t dW_t,
\end{align*}
\]

where \( r, \mu \) and \( \sigma \) are deterministic constants.

**Definition 2.3.3.** A stochastic variable \( \mathcal{X} \) is called a *contingent claim* if its value is determined by a stochastic process \((S_t)_{t \in [0,T]}\). In particular, if \( \mathcal{X} \) depends only on \( S_T \) (the value of the stochastic process at time \( T \)) then \( \mathcal{X} \) is called a *simple claim*. If \((S_t)_{t \in [0,T]}\) is a price process then \( \mathcal{X} \) is a *financial derivative*.

**Examples of simple contingent claims**

- European call option: At time \( T \), the holder of the option has the right, but not the obligation, to buy an asset (with price process \((S_t)_{t \in [0,T]}\)) at a predetermined strike price \( K \). This is a contingent claim that pays

\[ (S_T - K)^+ = \max(S_T - K, 0). \]

- European put option: At time \( T \), the holder of the option has the right, but not the obligation, to sell an asset (with price process \((S_t)_{t \in [0,T]}\)) at a predetermined strike price \( K \). This is a contingent claim that pays

\[ (K - S_T)^+ = \max(K - S_T, 0). \]

- European digital (call) option: At time \( T \), the holder of the option receives a fix amount \( D \) predetermined at the beginning of the contract if the price of the asset is greater than a predetermined strike price or zero otherwise. This contingent claim pays

\[
\begin{cases} 
D, & \text{if } S_T - K > 0, \\
0, & \text{otherwise.}
\end{cases}
\]

**Definition 2.3.4.** A portfolio \( \varphi = (\varphi^1, \cdots, \varphi^n) \) is a vector process. Each of its components describes the amount of a market security that we hold at time \( t \).

Consider a portfolio \((\phi, \psi)\) where \( \phi_t \) and \( \psi_t \) describe the number of units of one random security and the number of units of the bond that we hold at time \( t \) respectively. The processes can take positive or negative values
(we allow unlimited short-selling of the stock or bond). The security component of the portfolio $\phi$ should be $\mathcal{F}_t$-adapted.

A portfolio is self financing if and only if the change in its value only depends on the change of the asset prices. Let $S$ and $B$ denote the stock price and the bond price respectively, the value of a portfolio $(\phi_t, \psi_t)$ at time $t$ is given by $V_t = \phi_t S_t + \psi_t B_t$. At the next time instant, two things happen: the old portfolio changes value because $S_t$ and $B_t$ have changed price; and the old portfolio has to be adjusted to give a new portfolio as instructed by the trading strategy $(\phi, \psi)$. If the cost of the adjustment is perfectly matched by the profits or losses made by the portfolio then no extra money is required from outside, the portfolio is self financing.

In discrete time we get a difference equation.

$$
\Delta V_t = \phi_t \Delta S_t + \psi_t \Delta B_t
$$

In continuous time, we get a stochastic differential equation.

**Definition 2.3.5** (Self financing property). If $(\phi, \psi)$ is a portfolio with stock price $S$ and bond price $B$, then

$$(\phi, \psi)$$ is self financing $\iff dV_t = \phi_t dS_t + \psi_t dB_t$$

**Definition 2.3.6.** A market is said to be arbitrage free if it is not possible to purchase a portfolio for which there is a risk-less profit to be earned with positive probability.

**Definition 2.3.7** (Replicating strategy). Suppose that we are in a market of a riskless bond $B$, a risky security $S$, and a claim $X$ on events up to time $T$. A replicating strategy for $X$ is a self financing portfolio $(\phi, \psi)$ such that $V_T = \phi_T S_T + \psi_T B_T = X$.

Replicating strategies are important because if we are able to find a replicating strategy for the claim $X$, then the price of $X$ at time $t$ must be $V_t = \phi_t S_t + \psi_t B_t$ otherwise arbitrage opportunities will arise.

To obtain the replicating strategy for a given claim $X$, the first step is to use Theorem 7 (Girsanov Theorem) (take $\theta = \mu - r\sigma^2$) to find a measure $Q$ equivalent to $P$ under which the discounted stock process $Z_t = B_t^{-1} S_t$ is a martingale. (Note that because $\mu$, $\sigma$, and $r$ are deterministic constants the Novikov condition is satisfied).

Second, form the process $E_t = \mathbb{E}_Q [B_T^{-1} X | \mathcal{F}_t]$, this process is a martingale with respect to $Q$, note that from Proposition 2 (Tower Law) $\mathbb{E}_Q [\mathbb{E}_Q [X | \mathcal{F}_s] | \mathcal{F}_t] = \mathbb{E}_Q [X | \mathcal{F}_s]$, for $s \leq t$. Therefore, using Theorem 3 (Martingale Representation Theorem), it is possible to find an $\mathcal{F}_t$-adapted process $\phi$ such that $dE_t = \phi dZ_t$.

Consider the following replication strategy:

- hold $\phi_t$ units of stock at time $t$
- hold $\psi_t = E_t - \phi_t S_t$ units of the bond at time $t$.

The portfolio $(\phi, \psi)$ satisfies the self financing condition

$$dV_t = \phi_t dS_t + \psi_t dB_t$$

Note that $V_t = B_t E_t$, in particular the terminal value of the strategy is $V_T = E_T = X$, therefore there is an arbitrage price for $X$ at all times.
Theorem 8 (Risk Neutral Valuation Formula). Suppose that we have a Black-Scholes model for a continuously tradable stock and bond. Assume that the market is complete and arbitrage free. Let \( r \) be the (constant and deterministic) risk free interest rate and \( X \) be a claim, knowable by some time horizon \( T \). The arbitrage price of such claim \( X \) is given by

\[
V_t = B_t \mathbb{E}_Q \left[ B_T^{-1} X | F_t \right] = e^{-r(T-t)} \mathbb{E}_Q \left[ X | F_t \right]
\]

where \( Q \) is the martingale measure for the discounted stock \( B_t^{-1} S_t \).


Let \( X \) be a simple claim i.e. \( X = P(S_T) \) for some payoff function \( P \), from Theorem 7 (Girsanov Theorem) we obtain that \( W_t^Q = \mu - r \sigma t + W_t \), therefore the SDE of the process \( S_t \) written in terms of \( Q \) is

\[
dS_u = rS_u du + \sigma S_u dW_u^Q, \quad 0 \leq u < T,
\]

\[
S_t = s.
\]

Therefore

\[
S_T = s \exp \left( \left( r - \frac{1}{2} \sigma^2 \right) (T - t) + \sigma (W_T - W_t) \right).
\]

Let \( Y = \frac{\ln(S_T)}{s} \), we have that \( Y = \left( r - \frac{1}{2} \sigma^2 \right) (T - t) + \sigma (W_T - W_t) \), thus \( Y \) is normally distributed with mean \( \left( r - \frac{1}{2} \sigma^2 \right) (T - t) \) and standard deviation \( \sigma \sqrt{T - t} \).

However \( S_T = se^Y \), therefore, we obtain the pricing formula

\[
V(s,t) = e^{-r(T-t)} \int_{\mathbb{R}} P(se^y)p_m,\tilde{\sigma}(y)dy,
\]

where \( p_m,\tilde{\sigma} \) is the density function of a normally distributed random variable with mean \( m = \left( r - \frac{1}{2} \sigma^2 \right) (T - t) \) and standard deviation \( \tilde{\sigma} = \sigma \sqrt{T - t} \).

The previous result can be derived using another approach. Let \( V(S_t, t) \) denote the price of a simple claim \( X = P(S_T) \) at time \( t \), we assume as before that the market has a risky asset and risk free bond with dynamics:

\[
dB_t = rB_t dt,
\]

\[
dS_t = \mu S_t dt + \sigma S_t dW_t.
\]

where \( r, \mu \) and \( \sigma \) are deterministic constants.

Using Theorem 4 (Itô Formula) we obtain an equation for the infinitesimal change in the value of the claim:

\[
dV_t = d(V(S_t, t)) = \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S_t} dS_t + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2} dt
\]

Consider a portfolio consisting of 1 unit of the option and \(-\Delta \) units of the stock; at time \( t \) the value of this portfolio is

\[
\Pi_t = V(S_t, t) - \Delta S_t
\]

After an infinitesimal time increment of \( dt \), the change in portfolio value is given by

\[
d\Pi_t = \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S_t} dS_t + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V}{\partial S_t^2} dt - \Delta dS_t.
\]
Note that the right hand side of this equation has a deterministic and a random component. Therefore, choosing \( \Delta = \frac{\partial V}{\partial s} \) the randomness is reduced to zero. This choice of \( \Delta \) results in a risk free increment of the portfolio

\[
d\Pi_t = \left( \frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V}{\partial s^2} \right) dt.
\]

As a consequence, we have that

\[
d\Pi_t = r \Pi dt = r \left( V - \frac{\partial V}{\partial S} S_t \right) dt,
\]

otherwise arbitrage opportunities will arise.

Combining the two equations we obtain the **Black-Scholes equation**:

\[
\frac{\partial V}{\partial t} + rs \frac{\partial V}{\partial s} + \frac{1}{2} \sigma^2 s^2 \frac{\partial^2 V}{\partial s^2} - rV = 0.
\]

**Theorem 9.** The solution to the Black-Scholes equation:

\[
\frac{\partial V}{\partial t} + rs \frac{\partial V}{\partial s} + \frac{1}{2} \sigma^2 s^2 \frac{\partial^2 V}{\partial s^2} - rV = 0
\]

for \((t, s) \in [0, T) \times \mathbb{R}_+\)

with final condition \( V(T, s) = P(s) \) where \( P \in \{ f \in L^{1}_{\text{loc}} : f = O(|s|^{-\alpha} e^{s^2}), \alpha > 1 \} \) is

\[
V(s, t) = \frac{e^{-r(T-t)}}{\sigma \sqrt{2\pi (T-t)}} \int_{\mathbb{R}} P(se^{x}) \exp \left( \frac{(x - (r - \frac{1}{2} \sigma^2)(T-t))^2}{2\sigma^2(T-t)} \right) dx.
\]


Note that the price of a simple claim can be obtained either as a solution of a PDE or as expected value, this is always the case when the parameters \( r, \mu \) and \( \sigma \) are deterministic constants. The following theorem connects both results.

**Theorem 10** (Feynman-Kac formula). Suppose that \( X \) solves the stochastic differential equation:

\[
dX_s = f(X_s, s)ds + g(X_s, s)dW_s,
\]

where \( W \) is a 1-dimensional Brownian motion. Let \( b \) be a lower bounded continuous function defined on \([0, T] \times \mathbb{R}\), then the discounted expected payoff

\[
u(x, t) = \mathbb{E}_{X_t=x} \left[ e^{-\int_{t}^{T} b(X_s)ds} P(X_T) \right]
\]

solves the PDE

\[
u_t + f(x, t)u_x + \frac{1}{2} g^2(x, t)u_{xx} - b(x, t)u = 0, \quad \text{for } t < T, \quad \text{with } u(x, T) = P(x).
\]

**Proof.** See Oksendal, B. (2003), pp. 143-144.

### 2.3.2 The Multivariate Black-Scholes model

Let’s consider a market with \( n \) risky assets and a risk free bond. This model allows for correlation between the assets.

**Definition 2.3.8.** The Multivariate Black-Scholes model assumes a market of \( n + 1 \) assets with dynamics given
by
\[ dB_t = rB_t dt, \]
\[ dS_i^t = \mu_i S_i^t dt + \sigma_i S_i^t d\tilde{W}_i^t \quad \text{for } i = 1, \ldots, n. \]

where \( \tilde{W} = (\tilde{W}^1, \ldots, \tilde{W}^n) \) is vector of correlated Brownian motion.

From Definition 2.2.14 we have that \( \tilde{W} = \delta W \) where \( W \) is a vector of \( n \) (independent) Brownian motion and \( \rho = \delta \delta' \) is the correlation matrix. Therefore the market dynamics can be rewritten as
\[ dB_t = rB_t dt, \]
\[ dS_i^t = S_i^t (\mu_i dt + \sum_{j=1}^{n} \sigma_{ij} dW_j^t) \quad \text{for } i = 1, \ldots, n. \]

where \( [\sigma_{ij}] = M \delta \) with
\[
M = \begin{pmatrix}
\sigma^1 & 0 & \cdots & 0 \\
0 & \sigma^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \sigma^n
\end{pmatrix}
\]

**Definition 2.3.9** (Self financing property). Let \((\phi^1, \ldots, \phi^n, \psi)\) be a portfolio with risky securities \(S^i\) for \(i = 1, \ldots, n\) and risk free bond price \(B\), then
\[
(\phi^1, \ldots, \phi^n, \psi) \text{ is self financing } \iff \quad dV_t = \sum_{i=1}^{n} \phi_i^t dS_i^t + \psi_t dB_t,
\]
where \(V_t\) denotes the value of the portfolio at time \(t\).

**Definition 2.3.10.** A stochastic variable \(X\) is called a *contingent claim* if its value is determined by a stochastic vector process \(S = (S_t)_{t \in [0,T]}\) i.e., for \(t \in [0,T]\), \(S_t = (S_1^t, \ldots, S_n^t)\) where \(S^i\) is a stochastic process for \(i = 1, \ldots, n\). In particular, if \(X\) depends only on \(S_T\) then \(X\) is called a *simple claim*.

**Definition 2.3.11** (Replicating strategy). Suppose that we have a market of \(n\) risky securities \((S_1^t, \ldots, S^n_t)\), a riskless bond \(B\), and a claim \(X\) on events up to time \(T\). A *replicating strategy* for \(X\) is a self financing portfolio \((\phi^1, \ldots, \phi^n, \psi)\) such that \(V_T = \sum_{i=1}^{n} \phi_i^T S_T^i + \psi_T B_T = X\).

**Examples of Simple Claims:**

- **European Basket call option with weights \((w_i)_{i=1,\ldots,n}\) and strike price \(K\):** At time \(T\) the holder of the option receives the amount
\[
\left( \sum_{i=1}^{n} w_i S_i^T - K \right)^+ = \max \left( \sum_{i=1}^{n} w_i S_i^T - K, 0 \right),
\]
where \(\sum_{i=1}^{n} w_i = 1\) and \(w_i > 0\) for all \(i = 1, \ldots, n\).

- **Product option with strike price \(K\):** At time \(T\) the holder of the option receives the amount
\[
\left( \prod_{i=1}^{n} S_i^T - K \right)^+ = \max \left( \prod_{i=1}^{n} S_i^T - K, 0 \right).
• Portfolio of \( n \) European call options (with the same maturity): At time \( T \) the holder of the option receives

\[
\sum_{i=1}^{n} (S_T^i - K_i)^+ = \sum_{i=1}^{n} \max(S_T^i - K_i, 0).
\]

To derive the pricing formula for a simple claim \( \mathcal{X} \), we proceed as before. First, assume that \( \Sigma = [\sigma_{ij}] \) is invertible and let \( \mathbf{1} \) be the constant vector \((1, \cdots, 1)\). Note that \( \Sigma \theta = \mu - r \mathbf{1} \) has a unique solution: \( \theta = \Sigma^{-1}(\mu - r \mathbf{1}) \).

Let \( Z = (Z^1, \cdots, Z^n) \), using Theorem 7 (Girsanov Theorem) we obtain a measure \( \mathbb{Q} \) such that \( Z_t^i = B_t^{-1} S_t^i \) for \( i = 1, \cdots, n \) is a \( \mathbb{Q} \)-martingale. Let \( \mathcal{X} \) be a claim maturing at time \( T \) and define the process \( E_t = \mathbb{E}_\mathbb{Q}(B_T^{-1} \mathcal{X}|\mathcal{F}_t) \). This process is a \( \mathbb{Q} \)-martingale, therefore using Theorem 3 (Martingale Representation Theorem (\( n \)-dimensional version)), if the matrix \( \Sigma \) is invertible, then there exists a vector process \( \phi \) with \( \phi_t = (\phi_t^1, \cdots, \phi_t^n) \) such that

\[
E_t = E_0 + \sum_{i=1}^{n} \int_0^t \phi_s^i dZ_s^i.
\]

The hedging strategy \((\phi_t^1, \cdots, \phi_t^n, \psi_t)\) where \( \phi_t^i \) is the holding of asset \( i \) and \( \psi_t \) is the holding of the risk free bond at time \( t \), with \( \psi_t = E_t - \sum_{i=1}^{n} \phi_t^i Z_t^i \) guarantees that the portfolio \((\phi^1, \cdots, \phi^n, \psi)\) is self financing:

\[
dV_t = \sum_{i=1}^{n} \phi_t^i dS_t^i + \psi_t dB_t,
\]

also \( V_t = B_t E_t \) and the claim \( \mathcal{X} \) is attainable by this portfolio. Therefore, we have the following result for a market with \( n \) correlated risky assets and a risk free bond.

**Theorem 11** (Multivariate Risk Neutral Valuation Formula). Suppose that we have a market with \( n \) correlated risky assets and a risk free bond, with dynamics given by the Multivariate Black-Scholes model. Assume that the market is complete and arbitrage free. Let \( r \) be the (constant and deterministic) risk free interest rate and \( \mathcal{X} \) be a claim, knowable by some time horizon \( T \). The arbitrage price of such claim \( \mathcal{X} \) is given by

\[
V_t = B_t \mathbb{E}_\mathbb{Q} \left[ B_T^{-1} \mathcal{X}|\mathcal{F}_t \right] = e^{-r(T-t)} \mathbb{E}_\mathbb{Q} \left[ \mathcal{X}|\mathcal{F}_t \right],
\]

where \( \mathbb{Q} \) is the martingale measure for the discounted stock vector \( B_t^{-1} S_t \).

**Proof.** See Baxter, M. and Rennie, A. (1996), pp. 186-188. \(\square\)

A Black-Scholes pricing equation can be derived in the multidimensional setting too. Set up a portfolio of \( 1 \) unit of the financial derivative that we want to price and \(-\Delta_i \) units of each asset \( S^i \). At time \( t \) the value of this portfolio is

\[
\Pi_t = V(S_t^1, \cdots, S_t^n, t) - \sum_{i=1}^{n} \Delta_i S_t^i.
\]

Therefore, using Theorem 5 (General Itô Formula), we obtain that the change in value of this portfolio is given by

\[
d\Pi = \left( \frac{\partial V}{\partial t} + \frac{1}{2} \sum_{i,j=1}^{n} \sigma_i \sigma_j \rho_{ij} S_t^i S_t^j \frac{\partial^2 V}{\partial s_i \partial s_j} \right) dt + \sum_{i=1}^{n} \left( \frac{\partial V}{\partial s_i} - \Delta_i \right) dS_t^i.
\]

Choosing \( \Delta_i = \frac{\partial V}{\partial s_i} \), we obtain a risk free portfolio. In order to avoid arbitrage opportunities \( d\Pi_t = r\Pi dt \), which
leads to the following equation:

\[
\frac{\partial V}{\partial t} + \frac{1}{2} \sum_{i,j=1}^{n} \sigma_i \sigma_j \rho_{ij} s_i s_j \frac{\partial^2 V}{\partial s_i \partial s_j} + r \sum_{i=1}^{n} s_i \frac{\partial V}{\partial s_i} - rV = 0.
\]

This is the multidimensional version of the Black-Scholes equation.

**Theorem 12.** The solution to the Multidimensional Black-Scholes equation

\[
\frac{\partial V}{\partial t} + \frac{1}{2} \sum_{i,j=1}^{n} \sigma_i \sigma_j \rho_{ij} s_i s_j \frac{\partial^2 V}{\partial s_i \partial s_j} + r \sum_{i=1}^{n} s_i \frac{\partial V}{\partial s_i} - rV = 0
\]

with final condition

\[
V(T, s_1, \cdots, s_n) = P(s_1, \cdots, s_n)
\]

where \( P \in \{ f \in L^1_{loc} : f = O(|s|^{-\alpha} e^{-|s|^2}), \alpha > n \} \) is

\[
V(t, s) = \frac{e^{-r(T-t)}}{(2\pi)^{n/2} |\det(A)|^{1/2}} \int_{\mathbb{R}^n} P(s_1 e^{x_1}, \cdots, s_n e^{x_n}) \exp \left( -\frac{1}{2} (x-m)^T A^{-1} (x-m) \right) dx
\]

where

- \( A = (T-t)M[\rho_{ij}]M \) with (correlation matrix) \( [\rho_{ij}] = \delta \delta' \) and \( M \) a diagonal matrix with the \( \sigma_i \)'s in the main diagonal.
- \( m = (m_1, \cdots, m_n) \) with \( m_i = \left( r - \frac{\sigma_i^2}{2} \right) (T-t) \).


**Theorem 13** (General Feynman Kac formula). Suppose that \( X \) solves the vector-valued stochastic equation:

\[
dX^i_s = f_i(X,s)ds + \sum_j g_{ij}(X,s)dW^j_s,
\]

where each component of \( W \) is an independent Brownian motion. Let \( b \) be a lower bounded continuous function defined on \([0,T] \times \mathbb{R}^n\), then the discounted expected payoff

\[
u(x,t) = \mathbb{E}_{X_t=x} \left[ e^{-\int_t^T b(X,s)ds} P(X_T) \right]
\]

solves the PDE

\[
u_t + \mathcal{L}u - bu = 0, \text{ for } t < T, \text{ with } u(x, T) = P(x).
\]

where \( \mathcal{L} \) is the differential operator

\[
\mathcal{L}u = \sum_i f_i \frac{\partial u}{\partial x_i} + \frac{1}{2} \sum_{i,j,k} g_{ik} g_{jk} \frac{\partial^2 u}{\partial x_i \partial x_j}.
\]

**Proof.** See Oksendal, B. (2003), pp. 143-144.
2.4 Upper bounds

We want to obtain upper bounds on the magnitude of the price change of portfolios of single asset options after eigenvalues of the covariance matrix are modified i.e. we want to find upper bounds on

$$|u_A(s_t, t) - u_{\tilde{A}}(s_t, t)|,$$

where $u_A$ denotes the value of the portfolio and $A$ is the covariance matrix of the underlying assets and $\tilde{A}$ is obtained by modifying the eigenvalues of $A$ (the eigenvectors are kept constant). Our first approach to this problem was to consider the value of a multi-asset option as a function of the vector of eigenvalues of the covariance matrix. Note that a portfolio of several 1-asset options can be considered as a multi-asset option with payoff given by the sum of the payoffs of the 1-asset options, assuming that all the options in the portfolio have the same expiry date. We show the differentiability of the price of a multi-asset option as a function of the vector of eigenvalues, and use partial derivatives to obtain an upper bound when the covariance matrix is positive definite and eigenvalues are modified without making them zero. In a second approach to this problem, an upper bound is obtained when the eigenvalues are modified without making them zero, under the assumption that the covariance matrix is positive definite and the payoff is bounded, using an estimate of the $L^1$ distance between two non-degenerate Gaussian densities. Note that portfolios of digital or put options have bounded payoffs. Finally we consider the case when some eigenvalues are made zero. We start with the simplest case: portfolios of $n$ single asset options with uncorrelated underlying assets. Subsequently, portfolios of two single asset options with correlated underlying assets are considered and upper bounds are obtained when one of the eigenvalues is made zero.

Before we continue it is necessary to introduce some notation that will be used throughout this section.

**Notation:**

- $p_{A,m} : \mathbb{R}^n \to \mathbb{R}$ denotes a Gaussian density with mean vector $m$ and covariance matrix $A$.
- $D_\lambda$ is a diagonal matrix whose diagonal entries are the components of the vector $\lambda$.
- $A'$ is the transpose of matrix $A$.
- $t$ denotes time.
- $T$ denotes maturity of the contract.
- $\mathbb{I}(\cdot)$ is an indicator function
- $\delta(\cdot)$ denotes the Dirac delta function.
- $\Phi$ denotes the standard normal cumulative distribution function:
  $$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz$$
  for all $x \in \mathbb{R}$.
- The $Q$-function is the tail probability of a standard normal distribution:
  $$Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\frac{z^2}{2}} dz = 1 - \Phi(x)$$
for all \( x \in \mathbb{R} \).

### 2.4.1 Non-degenerate Gaussian densities. Mean value theorem.

**Theorem 14** (Mean Value). Let \( F : [c, d] \subset \mathbb{R} \rightarrow \mathbb{R} \) be continuous function on the interval \([c, d]\) and differentiable on \((c, d)\), then there exists \( \xi \in (c, d) \) such that:

\[
F(d) - F(c) = F'(\xi)(d - c).
\]


**Theorem 15** (Dominated Convergence). Let \((f_n)_n \) be a sequence in \( L^1(\mathbb{R}^n) \) such that \((a) f_n \rightarrow f \) a.e., and \((b)\) there exists a nonnegative function \( g \in L^1(\mathbb{R}^n) \) such that \(|f_n| \leq g \) a.e. for all \( n \). Then \( f \in L^1(\mathbb{R}^n) \) and 

\[
\int f = \lim_{n \rightarrow \infty} \int f_n.
\]


**Theorem 16** (Multidimensional Mean Value). Let \( U \subset \mathbb{R}^n \) be an open convex set and \( F : U \subset \mathbb{R}^n \rightarrow \mathbb{R} \) be differentiable at each \( x \in U \). For all \( a, b \in U \) there exists \( \theta \in [a, b] \) such that

\[
F(b) - F(a) = \sum_{j=1}^{n} \partial F/\partial x_j(\theta)(b_j - a_j),
\]

where \([a, b]\) is the segment connecting \( a \) and \( b \).


**Theorem 17** (Tonelli-Fubini). Suppose that \((X, M, \mu)\) and \((Y, N, \nu)\) are \( \sigma \)-finite measure spaces and let \( L^+(X \times Y) \) be the space of all measurable functions from \((X \times Y)\) to \([0, \infty]\).

(Tonelli) If \( f \in L^+(X \times Y) \), then the functions \( g(x) = \int f_x d\nu \) and \( h(y) = \int f^y d\mu \) are in \( L^+(X) \) and \( L^+(Y) \), respectively, and

\[
\int f d(\mu \times \nu) = \int \left[ \int f(x, y) d\nu(y) \right] d\mu(x) = \int \left[ \int f(x, y) d\mu(x) \right] d\nu(y). \quad (2.2)
\]

(Fubini) If \( f \in L^1(\mu \times \nu) \), then \( f_x \in L^1(\nu) \) for a.e. \( x \in X \), \( f^y \in L^1(\mu) \) for a.e. \( y \in Y \), the a.e.-defined functions \( g(x) = \int f_x d\nu \) and \( h(y) = \int f^y d\mu \) are in \( L^1(\mu) \) and \( L^1(\nu) \) respectively, and 2.2 holds.


Let \( A \) be a symmetric positive definite matrix, then \( A \) can be factored as

\[
A = QD_aQ^t,
\]

where \( Q \) is an orthogonal matrix, \( D_a \) is a diagonal matrix with diagonal entries \( a_j > 0 \) and \( a \) is a vector that contains the (positive) eigenvalues of \( A \).

Define

\[
B = QD_bQ^t,
\]
where \( Q \) is the orthogonal matrix previously obtained from the eigenvalue decomposition of \( A \), \( b \) is a vector, and \( D_b \) is a diagonal matrix with diagonal entries \( b_j > 0 \).

Let \( u_A(s_t, t) \) denote the price at time \( t \) of an option with payoff \( P(s_T) \) and covariance matrix \( A \), then

\[
 u_A(s_t, t) = \frac{e^{-r(T-t)}}{(2\pi)^{n/2} |\text{det}(A)|^{1/2}} \int_{\mathbb{R}^n} P(s_t e^x) \exp\left(-\frac{1}{2}(x - m)'A^{-1}(x - m)\right) dx
\]

where

\[
 m = (m_1, \cdots, m_n) \\
 m_j = r - \frac{\sigma_j^2}{2}(T-t) \\
 A = (T-t)D_\sigma C D_\sigma \\
 s_t e^x = (s_1^t e^{x_1}, \cdots, s_n^t e^{x_n})
\]

\( D_\sigma \) is a matrix with \( \sigma_j \) along the diagonal and zeros everywhere else, \( C \) is a correlation matrix, \( C = (\rho_{ij})_{ij} \) and \( r \) is the risk free interest rate.

Making the substitution \( x = Qy \) we obtain that

\[
 u_A(s_t, t) = \frac{e^{-r(T-t)}}{(2\pi)^{n/2} |\text{det}(A)|^{1/2}} \int_{\mathbb{R}^n} P(s_t e^{Qy}) \exp\left(-\frac{1}{2}(y - \tilde{m})'D_n^{-1}(y - \tilde{m})\right) dy
\]

\[
 = \int_{\mathbb{R}^n} e^{-r(T-t)} P(s_t e^{Qy}) p_{D_n}(\tilde{m})dy
\]

with \( \tilde{m} = Q'm \).

We consider \( u_A \) as a function of the vector of eigenvalues from the covariance matrix (keeping the rest of the parameters fixed including the mean vector \( \tilde{m} \) and the rotation matrix \( Q \)) i.e.

\[
 u(\lambda) = \int_{\mathbb{R}^n} e^{-r(T-t)} P(s_t e^{Qy}) p_{D_n}(\tilde{m})dy.
\]

Let \( a = (a_1, \cdots, a_n) \in \mathbb{R}^n \) with \( a_j > 0 \) for all \( j = 1, \cdots, n \), we show that there exists \( R > 0 \) such that \( \frac{\partial u}{\partial \lambda_j} \) exists and it is continuous on the ball \( B(a, R) \) for all \( j = 1, \cdots, n \).

Therefore, as a consequence of Theorem 16 (Multivariate Mean Value) we obtain that

\[
 |u_A(s_t, t) - u_B(s_t, t)| = |u(a) - u(b)| \\
 \leq C \sum_{j=1}^n |a_j - b_j|
\]

with

\[
 C = \max_j \max_{\lambda \in \overline{B(a,R)}} \left| \frac{\partial u}{\partial \lambda_j} (\lambda) \right| \\
 B = \frac{Q D_0 Q'}{R}
\]

for all \( b \in B(a, R) \).

More formally, we show that

**Proposition 18.** Let \( u_A(s_t, t) \) be the price at time \( t \) of an option with payoff function \( P(s_T) \) and positive definite covariance matrix \( A \). For any of the following payoff functions

\[
P(s_T) = \begin{cases} 
\sum_{i=1}^n \max(K_i - s^i_T, 0), & \text{portfolio of put options.} \\
\sum_{i=1}^n D_i \mathbb{I}(s^i_T - K_i > 0), & \text{portfolio of digital options.}
\end{cases}
\]
there exist $R, C > 0$ such that for all $b \in B(a, R)$,
\[
|u_A(s_t, t) - u_B(s_t, t)| \leq C \sum_{j=1}^{n} |a_j - b_j|,
\]
where $a = (a_1, \ldots, a_n)$ is the vector of eigenvalues of $A$ and $B$ is obtained from the eigenvalue decomposition of $A$ ($A = QD_0Q'$), by modifying (only) the vector of eigenvalues i.e. $B = QD_0Q'$, with $b = (b_1, \ldots, b_n)$.

**Proof.** Let $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ with $a_j > 0$ for all $j = 1, \ldots, n$. We show that:

1. $\frac{\partial u}{\partial \lambda_j}(\lambda)$ exists at $\lambda = a$ for all $j = 1, \ldots, n$.

2. There exists $R > 0$ such that $\frac{\partial u}{\partial \lambda_j}$ is continuous on $B(a, R)$ for all $j = 1, \ldots, n$.

Let $(a_j^{(k)}) \subset \mathbb{R}$ be a sequence that converges $a_j$ and define

\[
a_j^{(k)} = (a_1, \ldots, a_{j-1}, a_j^{(k)}, a_{j+1}, \ldots, a_n), \quad k \geq 1
\]

\[
g(y, \lambda) = e^{-\pi(T-t)} P(s_t e^Q y) P_{D_\lambda, \tilde{m}}(y)
\]

\[
h_k(y) = \frac{g(y, a^{(k)}) - g(y, a)}{a_j^{(k)} - a_j}.\]

We use Theorem 15 (Dominated Convergence) to show that

\[
\lim_{k \to \infty} \int_{\mathbb{R}^n} h_k(y)dy = \int_{\mathbb{R}^n} \lim_{k \to \infty} h_k(y)dy = \int_{\mathbb{R}^n} \frac{\partial u}{\partial \lambda_j}(y, a)dy < \infty.
\]

However, by definition:

\[
\frac{\partial u}{\partial \lambda_j}(a) = \lim_{k \to \infty} \frac{u(a^{(k)}) - u(a)}{a_j^{(k)} - a_j} = \lim_{k \to \infty} \int_{\mathbb{R}^n} h_k(y)dy.
\]

Therefore,

\[
\frac{\partial u}{\partial \lambda_j}(a) = \int_{\mathbb{R}^n} \frac{\partial u}{\partial \lambda_j}(y, a)dy < \infty.
\]

As $(a_j^{(k)})$ is convergent there exist $L, U > 0$ and $k_0 \in \mathbb{N}$ such that $L < a_j^{(k)} < U$ for all $k \geq k_0$.

Let

\[
a_\psi = (a_1, \ldots, a_{j-1}, \psi, a_{j+1}, \ldots, a_n)
\]

\[
a_L = (a_1, \ldots, a_{j-1}, L, a_{j+1}, \ldots, a_n)
\]

\[
a_U = (a_1, \ldots, a_{j-1}, U, a_{j+1}, \ldots, a_n)
\]

\[
p_\psi, \tilde{m}_j(y_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_j - \tilde{m}_j)^2}
\]

\[
p_L, \tilde{m}_j(y_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_j - \tilde{m}_j)^2}
\]

\[
p_U, \tilde{m}_j(y_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_j - \tilde{m}_j)^2}
\]

Note that,

\[
\frac{\partial g(y, \lambda)}{\partial \lambda_j} = \frac{1}{2} \left( \frac{(y_j - \tilde{m}_j)^2}{\lambda_j^2} - \frac{1}{\lambda_j} \right) g(y, \lambda)
\]
Next, we show that

\[ (1) \]

Hence, defining \( \psi \)

\[ (2) \]

Therefore, \( h \)

\[ (3) \]

with \( \psi \in [L, U] \)

\[ (4) \]

Also, there exists \( r > 0 \) such that for all \( \psi \in [L, U] \) and \( y_j > r \)

\[ (5) \]

Therefore,

\[ (6) \]

for all \( \psi \in [L, U] \) and \( y \notin B(\bar{m}, r) \subset \mathbb{R}^n \), where \( B(\bar{m}, r) \) is a closed ball centered at \( \bar{m} \) with radius \( r \).

Hence, defining

\[ (7) \]

with

\[ (8) \]

we obtain that

\[ (9) \]

with \( h \in L_1(\mathbb{R}^n) \).

From Theorem 15 (Dominated Convergence), we conclude that

\[ (10) \]

Note that what we just showed is that for any \( \lambda \in \mathbb{R}^n \) such that \( \lambda_i > 0 \) for all \( i = 1, \ldots, n \),

\[ (11) \]

Next, we show that \( \int_{\mathbb{R}^n} \frac{\partial g(y, \lambda)}{\partial \lambda_j} dy \) is continuous at any \( \lambda \in \mathbb{R}^n \) such that \( \lambda_i > 0 \) for all \( i = 1, \ldots, n \).
Let \( \lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n \) with \( \lambda_i > 0 \) for all \( i = 1, \ldots, n \) and \( (\lambda^{(k)}) \subset \mathbb{R}^n \) be a sequence that converges to \( \lambda \).

Then, there exist \( \epsilon > 0 \) and \( k_1 \in \mathbb{N} \) such that \( \lambda_i - \epsilon > 0 \) for all \( i = 1, \ldots, n \) and \( \lambda^{(k)} \in V \) for all \( k \geq k_1 \), where

\[
V = \{ \theta \in \mathbb{R}^n : \lambda_i - \epsilon \leq \theta_i \leq \lambda_i + \epsilon \}.
\]

Let

\[
p_{\theta_i, \bar{m}_i}(y_i) = \frac{1}{\sqrt{2\pi \theta_i}} e^{-\frac{1}{2} \left( \frac{y_i - \bar{m}_i}{\theta_i} \right)^2}
\]

\[
p_{\lambda_i + \epsilon, \bar{m}_i}(y_i) = \frac{1}{\sqrt{2\pi (\lambda_i + \epsilon)}} e^{-\frac{1}{2} \left( \frac{y_i - \bar{m}_i}{\lambda_i + \epsilon} \right)^2}
\]

\[
\lambda_{U} = (\lambda_1, \epsilon, \ldots, \lambda_n).
\]

For all \( i = 1, \ldots, n \) there exists \( r_i > 0 \) such that for all \( \theta = (\theta_1, \ldots, \theta_n) \in V \)

\[
p_{\theta_i, \bar{m}_i}(y) \leq p_{\lambda_i + \epsilon, \bar{m}_i}(y)
\]

for all \( y_i > r_i \). Therefore, taking \( r = \max(r_i) \) we obtain that

\[
p_{D_{\epsilon, \bar{m}}}(y) = \prod_{i=1}^{n} p_{\theta_i, \bar{m}_i}(y_i) 
\leq \prod_{i=1}^{n} p_{\lambda_i + \epsilon, \bar{m}_i}(y_i) = p_{D_{\lambda_{U}}}(y).
\]

Hence,

\[
\left| \frac{\partial}{\partial \lambda_j} g(y, \theta) \right| \leq \left| \frac{1}{2} e^{-r(T-t)} P(s_t e^{Qy}) \left( \frac{(y_j - \bar{m}_j)^2}{(\lambda_j - \epsilon)^2} + \frac{1}{\lambda_j - \epsilon} \right) p_{D_{\lambda_{U}}, \bar{m}}(y) \right|
\]

for all \( \theta \in V \) and \( y \notin B(\bar{m}, r) \).

Let

\[
h(y) = \begin{cases} 
M, & \text{if } y \in B(\bar{m}, r), \\
\frac{1}{2} e^{-r(T-t)} P(s_t e^{Qy}) \left( \frac{(y_j - \bar{m}_j)^2}{(\lambda_j - \epsilon)^2} + \frac{1}{\lambda_j - \epsilon} \right) p_{D_{\lambda_{U}}, \bar{m}}(y), & \text{if } y \notin B(\bar{m}, r), 
\end{cases}
\]

with

\[
M = \max_{y \in V, y \in B(\bar{m}, r)} \left( \frac{1}{2} e^{-r(T-t)} P(s_t e^{Qy}) \left( \frac{(y_j - \bar{m}_j)^2}{(\lambda_j - \epsilon)^2} - \frac{1}{\lambda_j} \right) p_{D_{\theta}, \bar{m}}(y) \right)
\]

and define

\[
f_k(y) = \frac{\partial}{\partial \lambda_j} g(y, \lambda^{(k)}).
\]

We have that \( f_k, h \in L^1(\mathbb{R}^n) \) and

\[
|f_k(y)| \leq \sup_{\theta \in V} \left| \frac{\partial}{\partial \lambda_j} g(y, \theta) \right| \leq h(y)
\]

for all \( k \geq k_1 \) and \( y \in \mathbb{R}^n \).

Therefore

\[
\lim_{k \to \infty} \frac{\partial}{\partial \lambda_j} (\lambda^{(k)}) = \lim_{k \to \infty} \int_{\mathbb{R}^n} f_k(y) dy = \int_{\mathbb{R}^n} \lim_{k \to \infty} f_k(y) dy = \int_{\mathbb{R}^n} \frac{\partial}{\partial \lambda_j} g(y, \lambda^{(k)}) dy = \int_{\mathbb{R}^n} \frac{\partial}{\partial \lambda_j} g(y, \lambda) dy = \frac{\partial}{\partial \lambda_j} (\lambda).
\]
As a consequence of Theorem 15 (Dominated Convergence) and the continuity of \( \frac{\partial u}{\partial \lambda_j} \).

Therefore, we conclude that for any \( \lambda = (\lambda_1, \cdots, \lambda_n) \in \mathbb{R}^n \) such that \( \lambda_i > 0 \) we have that \( \frac{\partial u}{\partial \lambda_i} \) is continuous at \( \lambda \). In particular, if \( A \) is a positive definite matrix, then \( A = QDQ' \), where \( a = (a_1, \cdots, a_n) \) with \( a_i > 0 \) for all \( i = 1, \cdots, n \). Hence, for each \( j = 1, \cdots, n \) there exists \( R_j > 0 \) such that \( \frac{\partial u}{\partial \lambda_j} \) is continuous on \( B(a, R_j) \). Taking \( R = \min_{j=1}^n R_j \), we obtain that \( \frac{\partial u}{\partial \lambda_j} \) is continuous on \( B(a, R) \) for all \( j = 1, \cdots, n \).

Therefore,

\[
|u_A(s_t, t) - u_B(s_t, t)| = |u(a) - u(b)| \leq \left( \max_j \frac{\partial u}{\partial \lambda_j} \right) \left( \sum_{j=1}^n |a_j - b_j| \right) \leq C \left( \sum_{j=1}^n |a_j - b_j| \right),
\]

with \( C = \max_j \max_{\lambda \in B(a, R)} \left| \frac{\partial u}{\partial \lambda_j}(\lambda) \right| \), as a consequence of Theorem 16 (Multivariate Mean Value).

2.4.2 \( L^1 \) distance inequalities

**Definition 2.4.1** (Hellinger distance). Let \( P \) and \( Q \) be two probability measures with densities \( p \) and \( q \) with respect to a dominating measure \( \lambda \). The Hellinger distance between two measures is defined as the \( L^2(\lambda) \) distance between the square roots of their densities:

\[
H(P, Q) = \left[ \int \left( \sqrt{\frac{dp}{d\lambda}} - \sqrt{\frac{dq}{d\lambda}} \right)^2 d\lambda \right]^{1/2} = \lambda[(\sqrt{p} - \sqrt{q})^2]^{1/2}.
\]

The Hellinger distance satisfies the inequality \( 0 \leq H(P, Q) \leq \sqrt{2} \). However some authors prefer to have an upper bound of 1, therefore they include an extra factor of \( \frac{1}{\sqrt{2}} \) in the definition of \( H(P, Q) \).

**Definition 2.4.2** (Hellinger affinity). Let \( P \) and \( Q \) be two probability measures with densities \( p \) and \( q \) with respect to a dominating measure \( \lambda \). The Hellinger affinity between two measures is defined as

\[
\alpha(P, Q) = \lambda[\sqrt{pq}] = \lambda\sqrt{pq}.
\]

**Definition 2.4.3** (Bhattacharya coefficient). Let \( p \) and \( q \) be probability densities with respect to the Lebesgue measure in \( \mathbb{R} \). The Bhattacharya coefficient is defined as

\[
BC(p, q) = \int_{\mathbb{R}} \sqrt{pq}.
\]

Note that if \( \lambda \) is the Lebesgue measure and \( P \) and \( Q \) are probability measures absolutely continuous with respect to \( \lambda \) then \( \alpha(P, Q) = BC(p, q) \).

We obtained an estimate of the change in price after modifying the eigenvalues of the covariance matrix based on some of the inequalities obtained by Kvatazde, Z. and Shervashidze, T. (2005).

**Proposition 19.** Let \( u_A(s_t, t) \) be the price at time \( t \) of an option with payoff \( P(s_T) \) and positive definite covari-
Corollary 19.1. Let \( u_A(s_t, t) \) be the price at time \( t \) of an option with positive definite covariance matrix \( A \) and payoff \( P(s_T) \) given by

\[
P(s_T) = \begin{cases} 
\sum_{i=1}^{n} \max(K_i - s_T^i, 0), & \text{portfolio of put options.} \\
\sum_{i=1}^{n} D_i \|s_T^i - K_i > 0\|, & \text{portfolio of digital options.}
\end{cases}
\]

then

\[
|u_A(s_t, t) - u_B(s_t, t)| \leq 2e^{-r(T-t)}||P||_\infty \sum_{i=1}^{n} \frac{|a_i - b_i|}{a_i + b_i},
\]

where \( a = (a_1, \ldots, a_n) \) is the vector of eigenvalues of \( A \) and \( B \) is obtained from the eigenvalue decomposition of \( A = QD_0Q' \), by modifying (only) the vector of eigenvalues i.e. \( B = QD_bQ' \), with \( b = (b_1, \ldots, b_n) \).

**Proof.** Let \( u_A(s_t, t) \) be the price at time \( t \) of an option with bounded payoff function \( P(s_T) \) and positive definite covariance matrix \( A \),

\[
u_A(s_t, t) = \frac{e^{-r(T-t)}}{(2\pi)^{\frac{n}{2}}|\det(A)|^{\frac{1}{2}}} \int_{\mathbb{R}^n} P(s_t e^x) \exp\left(-\frac{1}{2}(x - m)^T A^{-1} (x - m)\right) dx.
\]

Previously, we showed that by making the substitution \( x = Qy \) we obtain that

\[
u_A(s_t, t) = \frac{e^{-r(T-t)}}{(2\pi)^{\frac{n}{2}}|\prod_{i=1}^{n} \omega_i)^{\frac{1}{2}}} \int_{\mathbb{R}^n} P(s_t e^{Qy}) \exp\left(-\frac{1}{2}(y - \tilde{m})^T D_0^{-1} (y - \tilde{m})\right) dy
\]

with \( \tilde{m} = Q'm \).

As before, define \( B = QD_0Q' \) with \( b = (b_1, \cdots, b_n) \) such that \( b_i > 0 \) for all \( i = 1, \cdots, n \). Also, let \( p_{a_i, \tilde{m}_i} \) and \( p_{b_i, \tilde{m}_i} \) be Gaussian densities with mean \( \tilde{m}_i \) and variance \( a_i \) and \( b_i \) respectively for all \( i \), then

\[
p_{D_0, \tilde{m}} - p_{D_b, \tilde{m}} = \prod_{i=1}^{n} p_{a_i, \tilde{m}_i} - \prod_{i=1}^{n} p_{b_i, \tilde{m}_i} = p_{a_1, \tilde{m}_1} \cdots p_{a_n, \tilde{m}_n} - p_{b_1, \tilde{m}_1} \cdots p_{b_n, \tilde{m}_n}.
\]

Moreover,

\[
\int_{\mathbb{R}^n} p_{D_0, \tilde{m}} - p_{D_b, \tilde{m}} = \sum_{i=1}^{n} \int_{\mathbb{R}} p_{a_i, \tilde{m}_i} - p_{b_i, \tilde{m}_i}.
\]

This can be proved using mathematical induction:

Case \( k = 1 \) is trivial.

Case \( k = n \):

Assume that the result is true for \( k = n - 1 \):
\[
\int_{\mathbb{R}^{n-1}} p_{\alpha_1,\tilde{m}_1} \cdots p_{\alpha_{n-1},\tilde{m}_{n-1}} - p_{b_1,\tilde{m}_1} \cdots p_{b_{n-1},\tilde{m}_{n-1}} = \sum_{i=1}^{n-1} \int_{\mathbb{R}} p_{\alpha_i,\tilde{m}_i} - p_{b_i,\tilde{m}_i}.
\]

Then
\[
\int_{\mathbb{R}^n} p_{\alpha_1,\tilde{m}_1} \cdots p_{\alpha_n,\tilde{m}_n} - p_{b_1,\tilde{m}_1} \cdots p_{b_n,\tilde{m}_n} = \int_{\mathbb{R}^n} p_{\alpha_1,\tilde{m}_1} \cdots p_{\alpha_{n-1},\tilde{m}_{n-1}} p_{\alpha_n,\tilde{m}_n} \\
- \int_{\mathbb{R}^n} p_{\alpha_1,\tilde{m}_1} \cdots p_{\alpha_{n-1},\tilde{m}_{n-1}} p_{b_{n-1},\tilde{m}_{n-1}} \\
+ \int_{\mathbb{R}^n} p_{\alpha_1,\tilde{m}_1} \cdots p_{\alpha_{n-1},\tilde{m}_{n-1}} p_{b_n,\tilde{m}_n} \\
- \int_{\mathbb{R}^n} p_{b_1,\tilde{m}_1} \cdots p_{b_{n-1},\tilde{m}_{n-1}} p_{\alpha_n,\tilde{m}_n} \\
+ \int_{\mathbb{R}^n} p_{b_1,\tilde{m}_1} \cdots p_{b_{n-1},\tilde{m}_{n-1}} (p_{\alpha_n,\tilde{m}_n} - p_{b_n,\tilde{m}_n}) \\
- \int_{\mathbb{R}^n} p_{b_1,\tilde{m}_1} \cdots p_{b_{n-1},\tilde{m}_{n-1}} p_{b_n,\tilde{m}_n} \\
= \int_{\mathbb{R}^n} p_{\alpha_n,\tilde{m}_n} - p_{b_n,\tilde{m}_n} \\
\sum_{i=1}^{n-1} \int_{\mathbb{R}} p_{\alpha_i,\tilde{m}_i} - p_{b_i,\tilde{m}_i} \\
= \sum_{i=1}^{n-1} \int_{\mathbb{R}} p_{\alpha_i,\tilde{m}_i} - p_{b_i,\tilde{m}_i}.
\]

Therefore
\[
\int_{\mathbb{R}^n} |p_{\alpha_n,\tilde{m}_n} - p_{b_n,\tilde{m}_n}| \leq \sum_{i=1}^{n} \int_{\mathbb{R}} |p_{\alpha_i,\tilde{m}_i} - p_{b_i,\tilde{m}_i}|.
\]

In addition, let \( p \) and \( q \) be probability densities with respect to the Lebesgue measure in \( \mathbb{R} \). Using the Cauchy-Schwartz inequality we obtain that
\[
\int_{\mathbb{R}} |p - q| = \int_{\mathbb{R}} |\sqrt{p} - \sqrt{q}| |\sqrt{p} + \sqrt{q}| \\
\leq \left( \int_{\mathbb{R}} |\sqrt{p} - \sqrt{q}|^2 \right)^{\frac{1}{2}} \left( \int_{\mathbb{R}} |\sqrt{p} + \sqrt{q}|^2 \right)^{\frac{1}{2}} \\
= (2(1 - BC(p, q)))^{\frac{1}{2}} (2(1 + BC(p, q)))^{\frac{1}{2}} \\
= 2\sqrt{1 - BC^2(p, q)}.
\]

This inequality is part of the more general result:
\[
\int_{\mathbb{R}} |p - q| \leq 2H(P, Q) \leq 2 \left( \int_{\mathbb{R}} |p - q| \right)^{\frac{1}{2}}.
\]

Note that
\[
H^2(P, Q) = \int_{\mathbb{R}} (\sqrt{p} - \sqrt{q})^2 \\
= 2(1 - \int_{\mathbb{R}} \sqrt{pq}) \\
= 2(1 - BC(p, q)).
\]

Therefore
\[
\int_{\mathbb{R}} |p - q| \leq (2(1 - BC(p, q)))^{\frac{1}{2}} (2(1 + BC(p, q)))^{\frac{1}{2}} \\
= H(P, Q)\sqrt{4 - H^2(P, Q)} \\
\leq 2H(P, Q)
\]

because
\[
0 \leq H(P, Q) \leq \sqrt{2}.
\]
Similarly,
\[
(\sqrt{p} - \sqrt{q})^2 \leq |\sqrt{p} - \sqrt{q}| (\sqrt{p} + \sqrt{q}) = |p - q|.
\]
Therefore,
\[
H(\mathbb{P}, \mathbb{Q}) \leq \left( \int |p - q| \right)^{\frac{1}{2}}.
\]
Making \( p = p_{a_i, \tilde{m}_i} \) and \( q = p_{b_i, \tilde{m}_i} \), we obtain that
\[
BC(p_{a_i, \tilde{m}_i}, p_{b_i, \tilde{m}_i}) = \int_{\mathbb{R}} \sqrt{p_{a_i, \tilde{m}_i}(y_1)p_{b_i, \tilde{m}_i}(y_1)} dy_1
= \left( \frac{1}{2\pi\sqrt{a_i b_i}} \right)^{\frac{1}{2}} \int_{\mathbb{R}} \exp \left( -\frac{1}{2} \left( \left( \frac{1}{a_i} + \frac{1}{b_i} \right) \frac{1}{2} (y_1 - \tilde{m}_i)^2 \right) \right) dy_1
= \left( \frac{1}{2\pi\sqrt{a_i b_i}} \right)^{\frac{1}{2}} \int_{\mathbb{R}} \exp \left( -\frac{1}{2} \left( \frac{a_i + b_i}{2a_i b_i} \right) \frac{1}{2} (y_1 - \tilde{m}_i)^2 \right) dy_1
= \left( \frac{1}{2\pi\sqrt{a_i b_i}} \right)^{\frac{1}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz_i
= \left( \frac{1}{2\pi\sqrt{a_i b_i}} \right)^{\frac{1}{2}},
\]
after making the change of variables \( z_i = \left( \frac{a_i + b_i}{2a_i b_i} \right)^{\frac{1}{2}} (y_1 - \tilde{m}_i) \).
Therefore,
\[
\int_{\mathbb{R}} |p_{a_i, \tilde{m}_i} - p_{b_i, \tilde{m}_i}| \leq 2\sqrt{1 - BC^2(p_{a_i, \tilde{m}_i}, p_{b_i, \tilde{m}_i})}
= 2\sqrt{\frac{a_i - 2\sqrt{a_i b_i} + b_i}{a_i + b_i}}
= 2\sqrt{\frac{\sqrt{a_i} - \sqrt{b_i}}{\sqrt{a_i} + \sqrt{b_i}}}
= 2\sqrt{\frac{\sqrt{a_i} - \sqrt{b_i}}{\sqrt{a_i} + \sqrt{b_i}}} \cdot \sqrt{\frac{\sqrt{a_i} - \sqrt{b_i}}{\sqrt{a_i} + \sqrt{b_i}}}
= 2\left| \frac{a_i - b_i}{a_i + b_i} \right|
= 2\left( \frac{a_i - b_i}{a_i + b_i} \right)^2.
\]
Hence,
\[
|u_A(s_t, t) - u_B(s_t, t)| \leq \int_{\mathbb{R}^n} e^{-r(T-t)} P(s_t e^{Qy}) |p_{D_{a_i, \tilde{m}_i}}(y) - p_{D_{b_i, \tilde{m}_i}}(y)| dy
\leq e^{-r(T-t)} ||P||_{\infty} \sum_{i=1}^{n} \int_{\mathbb{R}} |p_{a_i, \tilde{m}_i} - p_{b_i, \tilde{m}_i}| \leq 2e^{-r(T-t)} ||P||_{\infty} \sum_{i=1}^{n} \frac{|a_i - b_i|}{a_i + b_i}.
\]
Monte Carlo simulations were used to compare the magnitude of the difference between the price of a portfolio of 2 put options and the approximation obtained after modifying the smallest eigenvalue, with our proposed upper bound. The following values were used: \( S_i^t = 10 \) and \( K_i = 9.6 \) for \( i = 1, 2, \Delta = T - t = 1, \sigma = 0.2, \sigma_2 = 0.3, \rho = 0.3 \) and \( r = 0.02 \). The total number of simulations used was 2000000. Figure 2.1 shows the results.

### 2.4.3 Uncorrelated underlying assets

**Theorem 20** (Markov’s Inequality). Let \( (\Omega, \mathcal{F}, \mathbb{P}) \) be a probability space and \( X : (\Omega, \mathcal{F}, \mathbb{P}) \to [0, \infty) \) be a random variable, then for any \( a > 0 \)
\[
\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}.
\]
Figure 2.1: Magnitude of the difference between the price of a portfolio of 2 put options and the approximation obtained after modifying the smallest eigenvalue without making it zero, and the proposed upper bound.

**Proposition 21** (Chernoff bound of the $Q$ function).

$$Q(x) \leq e^{-\frac{x^2}{2}} \text{ for all } x > 0.$$  

**Proof.** Let $X$ be a random variable with standard normal distribution and $x > 0$. For any $\lambda > 0$, we have that

$$Q(x) = P(X > x) = P(e^{\lambda X} > e^{\lambda x}) \leq e^{-\lambda x}E(e^{\lambda X})$$

from Markov’s inequality.

However,

$$E(e^{\lambda X}) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{\lambda x} e^{-\frac{1}{2}x^2} dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2 - 2\lambda x + \lambda^2) + \lambda^2} dx = e^{\frac{\lambda^2}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x - \lambda)^2} dx = e^{\frac{\lambda^2}{2}}.$$

Hence,

$$Q(x) \leq e^{-\lambda x} e^{\frac{\lambda^2}{2}}.$$

In particular, if we choose $\lambda = x$ we obtain that

$$Q(x) \leq e^{-\frac{x^2}{2}}.$$

**Proposition 22.** Let $U$ be a (Lebesgue) measurable subset of $\mathbb{R}$ then

$$\frac{1}{\sqrt{2\pi}a} \int_U e^{cx} e^{-\frac{(x-\mu)^2}{2a}} dx = \frac{1}{\sqrt{2\pi}c^2} \int_U e^{-\frac{(x-(\mu+c))^2}{2a}} dx.$$
Proof.
\[
\frac{1}{\sqrt{2\pi}} \int_U e^{cx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \frac{1}{\sqrt{2\pi}} \int_U e^{-\frac{(x^2 - 2x\mu + \mu^2)}{2\sigma^2}} \, dx
\]
\[
= \frac{1}{\sqrt{2\pi}} \int_U e^{-\frac{x^2 - 2x\mu + \mu^2 + 2\sigma^2}{2\sigma^2}} \, dx
\]
\[
= \frac{1}{\sqrt{2\pi}} \int_U e^{-\frac{x^2}{2\sigma^2} - \frac{2(\sigma^2 + \mu^2 - \mu^2)}{2\sigma^2}} \, dx
\]
\[
= \frac{1}{\sqrt{2\pi}} \int_U e^{-\frac{x^2 - 2(\sigma^2 + \mu^2 - \mu^2)}{2\sigma^2}} \, dx
\]
\[
= \frac{1}{\sqrt{2\pi}} \int_U e^{-\frac{x^2}{2\sigma^2} - \frac{2\sigma^2 + \mu^2 - \mu^2}{2\sigma^2}} \, dx
\]
\[
= \frac{1}{\sqrt{2\pi}} \int_U e^{-\frac{x^2}{2\sigma^2}} \, dx = \frac{1}{\sqrt{2\pi}} \int_U e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx
\]

Proposition 23. Let \( U \) be a (Lebesgue) measurable subset of \( \mathbb{R} \) then
\[
\int_U e^{-\frac{(cx)^2}{a}} e^{-\frac{(x-\mu)^2}{b}} \, dx = e^{-\frac{(\mu - cx)^2}{a+b}} \int_U e^{-\frac{(x-\mu)^2}{b}} \, dx.
\]

Proof. We have that
\[
\frac{(cx-\mu)^2}{b} - \frac{(x-\mu)^2}{a} = \frac{a(cx^2 - 2cx\mu + \mu^2) + b(x^2 - 2x\mu + \mu^2)}{ab}
\]
\[
= \frac{a(b^2 + \mu^2) - (b^2 + \mu^2)}{ab} \left( x^2 - 2 \left( \frac{ac\mu^2 + b^2}{ac^2 + b} \right) x + \left( \frac{ac\mu^2 + b^2}{ac^2 + b} \right)^2 \right) + \frac{a(b^2 + \mu^2)}{ab}
\]
\[
= \frac{a(b^2 + \mu^2)}{ab} \left( x - \left( \frac{ac\mu^2 + b^2}{ac^2 + b} \right) \right)^2 + \frac{1}{ab} \left( \frac{ac\mu^2 + b^2}{ac^2 + b} \right)^2 - \left( \frac{ac\mu^2 + b^2}{ac^2 + b} \right)^2.
\]

However,
\[
\frac{1}{ab} \left( \frac{(ac\mu^2 + b^2)}{ac^2 + b} \right)^2 - \left( \frac{ac\mu^2 + b^2}{ac^2 + b} \right)^2 = \frac{1}{ab} \left( \frac{a^2b^2 + 2abc\mu^2 + b^2\mu^2 - a^2c^2\mu^2 - ab\mu^2 + abc\mu^2 + b^2\mu^2}{ac^2 + b} \right)
\]
\[
= \frac{1}{ab} \left( \frac{2abc\mu^2 - abc\mu^2}{ac^2 + b} \right)
\]
\[
= \frac{(\mu^2 - 2\mu^2)}{ac^2 + b}
\]
\[
= \frac{(\mu^2 - 2\mu^2)}{ac^2 + b}.
\]

Hence,
\[
\int_U e^{-\frac{(cx)^2}{a}} e^{-\frac{(x-\mu)^2}{b}} \, dx = e^{-\frac{(\mu - cx)^2}{a+b}} \int_U e^{-\frac{(x-\mu)^2}{b}} \, dx.
\]

Let \( u_A(s, t) \) be the price of an option with payoff \( P(S_T) \) and positive definite covariance matrix \( A \) then
\[ u_A(s_t, t) = \frac{e^{-r(T-t)}}{(2\pi)^{\frac{n}{2}} |\det(A)|^{\frac{1}{2}}} \int_{\mathbb{R}^n} P(s_t e^x) \exp\left(-\frac{1}{2}(x-m)^T A^{-1} (x-m) \right) dx. \]

In particular, if the underlying assets are uncorrelated (\(A = D_a\)) then

\[ u_A(s_t, t) = \frac{e^{-r(T-t)}}{(2\pi)^{\frac{n}{2}} (\prod_{i=1}^a s_i)^{\frac{1}{2}}} \int_{\mathbb{R}^n} P(s_t e^x) \exp\left(-\frac{1}{2}(x-m)^T D_a^{-1} (x-m) \right) dx = \int_{\mathbb{R}^n} e^{-r(T-t)} P(s_t e^x) p_{D_a,m}(x) dx. \]

where \(D_a\) is a diagonal matrix and \(a = (a_1, \cdots, a_n)\) is the vector of diagonal entries of \(D_a\).

As a consequence, if \(P(S_T) = \sum_{i=1}^n D_i 1(S_T^i > K_i > 0)\) (portfolio of digital options) then

\[ u_A(s_t, t) = \int_{\mathbb{R}^n} e^{-r(T-t)} \left( \sum_{i=1}^n D_i 1(s_t e^x_i > K_i > 0) \right) p_{D_a,m}(x) dx = \int_{\mathbb{R}^n} e^{-r(T-t)} \sum_{i=1}^n D_i 1(s_t e^x_i > K_i > 0) \prod_{i=1}^n p_{a_i,m_i}(x_i) dx. \]

Similarly, if \(P(S_T) = \sum_{i=1}^n \max(S_T^i - K_i, 0)\) (portfolio of call options) then

\[ u_A(s_t, t) = \int_{\mathbb{R}^n} e^{-r(T-t)} \left( \sum_{i=1}^n \max \left( s_t e^x_i - K_i, 0 \right) \right) p_{D_a,m}(x) dx = \int_{\mathbb{R}^n} e^{-r(T-t)} \sum_{i=1}^n D_i 1(s_t e^x_i - K_i > 0) \prod_{i=1}^n p_{a_i,m_i}(x_i) dx. \]

Let \( \tilde{A} = D_a \) with \( \tilde{a} = (a_1, \cdots, a_k, 0, \cdots, 0) \) we show that

1. If \( u_A^D \) denotes the price of a portfolio of \( n \) digital options (with the same maturity) at time \( t \) and covariance matrix \( A \) then there exist constants \( C_{j1}, C_{j2} > 0 \) with \( j = 1, \cdots, n-k \) such that

\[ |u_A^D(s_t, t) - u_A^D(s_t, t)| \leq \sum_{j=1}^{n-k} C_{j1} e^{-\frac{C_{j2}}{x_{k+j}}}. \]

2. If \( u_A^C \) denotes the price of a portfolio of \( n \) call options (with the same maturity) at time \( t \) and covariance matrix \( A \) then there exist constants \( C_{j1}, C_{j2}, C_{j3}, C_{j4} > 0 \) with \( j = 1, \cdots, n-k \) such that

\[ |u_A^C(s_t, t) - u_A^C(s_t, t)| \leq \sum_{j=1}^{n-k} C_{j1} (e^{C_{j2} x_{k+j}} - 1) + C_{j3} e^{-\frac{C_{j4}}{x_{k+j}}}. \]

More formally, we have the following proposition:

**Proposition 24.** Let \( u_A^D(s_t, t) \) denote the price of a portfolio of \( n \) digital options (with the same maturity) at time \( t \) with positive definite covariance matrix \( A \). If the underlying assets are uncorrelated i.e. \( A = D_a\) then there exist constants \( C_{j1}, C_{j2} > 0 \) with \( j = 1, \cdots, n-k \) such that

\[ |u_A^D(s_t, t) - u_A^D(s_t, t)| \leq \sum_{j=1}^{n-k} C_{j1} e^{-\frac{C_{j2}}{x_{k+j}}}. \]

where \( u_A^D(s_t, t) \) denotes an approximation of the price obtained by making zero \( n - k \) eigenvalues of the covariance matrix i.e \( \tilde{A} = D_a \) with \( \tilde{a} = (a_1, \cdots, a_k, 0, \cdots, 0) \).
Proof. If the underlying assets are uncorrelated then $A = D_a$ where $a = (a_1, \cdots, a_n)$ with $a_i > 0$ for all $i = 1, \cdots, n$, therefore if $u^D_A(s_t, t)$ denotes the price of a portfolio of digital $n$ options at time $t$ and covariance matrix $A$ then

$$u^D_A(s_t, t) = e^{-r(T-t)} \sum_{i=1}^{n} \int_{\mathbb{R}} D_i \ln (s_t e^{x_i} - K_i > 0) p_{a_i, m_i}(x_i) dx_i$$

$$= e^{-r(T-t)} \sum_{i=1}^{n} \int_{\mathbb{R}} D_i p_{a_i, m_i}(x_i) dx_i.$$  

In addition, if $\tilde{A} = D_{\tilde{a}}$ with $\tilde{a} = (a_1, \cdots, a_k, 0, \cdots, 0)$ then

$$u^D_{\tilde{A}}(s_t, t) = e^{-r(T-t)} \sum_{i=1}^{k} \int_{\mathbb{R}} D_i p_{a_i, m_i}(x_i) dx_i + e^{-r(T-t)} \sum_{i=k+1}^{n} \int_{\mathbb{R}} D_i \delta(x_i - m_i) dx_i.$$  

Let

$$u_{a_i}(s_t^i, t) = \int_{\ln \frac{K_i}{s_t}}^{\infty} p_{a_i, m_i}(x_i) dx_i$$

and

$$u_0(s_t^i, t) = \int_{\ln \frac{K_i}{s_t}}^{\infty} \delta(x_i - m_i) dx_i,$$

we have that

$$|u^D_A(s_t, t) - u^D_{\tilde{A}}(s_t, t)| \leq e^{-r(T-t)} \sum_{i=k+1}^{n} D_i |u_{a_i}(s_t^i, t) - u_0(s_t^i, t)|.$$  

However,

$$u_{a_i}(s_t^i, t) = \int_{\ln \frac{K_i}{s_t}}^{\infty} \frac{1}{\sqrt{2\pi a_i}} e^{-\frac{(x_s - m_i)^2}{2a_i}} dx_i$$

$$= \int_{\ln \frac{K_i}{s_t}}^{\infty} \frac{1}{\sqrt{2\pi a_i}} e^{-\frac{(x_s'^i)^2}{2}} dx_i$$

$$= \left(1 - \Phi \left( \frac{\ln \frac{K_i}{s_t} - m_i}{\sqrt{a_i}} \right) \right)$$

$$= Q \left( \frac{\ln \frac{K_i}{s_t} - m_i}{\sqrt{a_i}} \right).$$

Also,

$$u_0(s_t^i, t) = \begin{cases} 1, & \text{if } m_i > \ln \frac{K_i}{s_t}, \\ \frac{1}{2}, & \text{if } m_i = \ln \frac{K_i}{s_t}, \\ 0, & \text{if } m_i < \ln \frac{K_i}{s_t}. \end{cases}$$

Note that if $\ln \frac{K_i}{s_t} = m_i$ then $|u_{a_i}(s_t^i, t) - u_0(s_t^i, t)| = 0$ for all $a_i > 0$. Therefore, only the other two cases are nontrivial:

1. $m_i > \ln \frac{K_i}{s_t}$
2. $m_i < \ln \frac{K_i}{s_t}$
Case 1: If \( m_i > \ln \frac{K_i}{s_i} \) then

\[
|u_{A_i}(s_i^t, t) - u_0(s_i^t, t)| = 1 - Q \left( \frac{\ln \frac{K_i}{s_i} - m_i}{\sqrt{a_i}} \right) \\
= Q \left( \frac{m_i - \ln \frac{K_i}{s_i}}{\sqrt{a_i}} \right) \\
\leq e^{-\frac{(m_i - \ln \frac{K_i}{s_i})^2}{2a_i}}.
\]

Case 2: If \( m_i < \ln \frac{K_i}{s_i} \) then

\[
|u_{A_i}(s_i^t, t) - u_0(s_i^t, t)| = Q \left( \frac{\ln \frac{K_i}{s_i} - m_i}{\sqrt{a_i}} \right) \\
\leq e^{-\frac{(m_i - \ln \frac{K_i}{s_i})^2}{2a_i}}.
\]

Remark. Note that both inequalities are obtained using Proposition 21 (Chernoff bound of the \( Q \) function).

Hence, we obtain that there exist constants \( C_1^j, C_2^j > 0 \) with \( j = 1, \cdots, n - k \) such that

\[
|u^{D_A}_i(s_t, t) - u^{D_{\hat{A}}}_i(s_t, t)| \leq \sum_{j=1}^{n-k} C_1^j e^{-\frac{C_2^j}{a_{k+j}}}.
\]

\( \square \)

A similar result holds for call options.

**Proposition 25.** Let \( u^c_A(s_t, t) \) denote the price of a portfolio of \( n \) call options (with the same maturity) at time \( t \) with positive definite covariance matrix \( A \). If the underlying assets are uncorrelated i.e. \( A = D_a \) then there exist constants \( C_1^j, C_2^j, C_3^j, C_4^j > 0 \) with \( j = 1, \cdots, n - k \) such that

\[
|u^{C_A}_i(s_t, t) - u^{C_{\hat{A}}}_i(s_t, t)| \leq \sum_{j=1}^{n-k} C_1^j (e^{C_2^j a_{k+j}} - 1) + C_4^j e^{-\frac{C_2^j}{a_{k+j}}},
\]

where \( u^c_A(s_t, t) \) denotes an approximation of the price obtained by making zero \( n-k \) eigenvalues of the covariance matrix i.e \( \hat{A} = D_{\hat{a}} \) with \( \hat{a} = (a_1, \cdots, a_k, 0, \cdots, 0) \).

**Proof.** If the underlying assets are uncorrelated then \( A = D_a \) where \( a = (a_1, \cdots, a_n) \) with \( a_i > 0 \) for all \( i = 1, \cdots, n \), therefore if \( u^{C_A}_i(s_t, t) \) denotes the price of a portfolio of digital options at time \( t \) and covariance matrix \( A \) then

\[
u^{C_A}_i(s_t, t) = e^{-r(T-t)} \sum_{i=1}^{n} \int_{\mathbb{R}} \max(s_i^t e^{x_i} - K_i, 0) p_{a_i, m_i}(x_i) dx_i.
\]

In addition, if \( \hat{A} = D_{\hat{a}} \) with \( \hat{a} = (a_1, \cdots, a_k, 0, \cdots, 0) \) then

\[
u^{C_{\hat{A}}}_i(s_t, t) = e^{-r(T-t)} \sum_{i=1}^{k} \int_{\mathbb{R}} \max(s_i^t e^{x_i} - K_i) p_{a_i, m_i}(x_i) dx_i + e^{-r(T-t)} \sum_{i=k+1}^{n} \int_{\mathbb{R}} \max(s_i^t e^{x_i} - K_i) \delta(x_i - m_i) dx_i.
\]
Let 

\[ u_{a_i}(s_i^t, t) = \int_{\ln \frac{K_i}{s_i^t}}^{\infty} (s_i^t e^{s_i^t} - K_i) p_{a_i, m_i}(x_i) \, dx_i \]

and 

\[ u_0(s_i^t, t) = \int_{\ln \frac{K_i}{s_i^t}}^{\infty} (s_i^t e^{s_i^t} - K_i) \delta(x_i - m_i) \, dx_i, \]

we have that 

\[ |u_A^C(s_i^t, t) - u_A^C(s_i^t, t)| \leq e^{-(T-t)} \sum_{i=1}^{n} |u_{a_i}(s_i^t, t) - u_0(s_i^t, t)|. \]

Using Proposition 22 we obtain that 

\[ \int_{\ln \frac{K_i}{s_i^t}}^{\infty} s_i^t e^{s_i^t} p_{a_i, m_i}(x_i) \, dx_i = \int_{\ln \frac{K_i}{s_i^t}}^{\infty} \frac{1}{\sqrt{2\pi a_i}} e^{-\frac{(x_i-m_i)^2}{2a_i}} s_i^t e^{s_i^t} \, dx_i \]

\[ = s_i^t e^{m_i + \frac{a_i}{2}} \int_{\ln \frac{K_i}{s_i^t}}^{\infty} \frac{1}{\sqrt{2\pi a_i}} e^{-\frac{(x_i-(m_i+a_i))^2}{2a_i}} \, dx_i' \]

\[ = s_i^t e^{m_i + \frac{a_i}{2}} \left( 1 - \Phi\left( \frac{\ln \frac{K_i}{s_i^t} - (m_i + a_i)}{\sqrt{a_i}} \right) \right) \]

\[ = s_i^t e^{m_i + \frac{a_i}{2}} Q\left( \frac{\ln \frac{K_i}{s_i^t} - (m_i + a_i)}{\sqrt{a_i}} \right). \]

On the other hand,

\[ K_i \int_{\ln \frac{K_i}{s_i^t}}^{\infty} \frac{1}{\sqrt{2\pi a_i}} e^{-\frac{(x_i-m_i)^2}{2a_i}} \, dx_i = K_i \int_{\ln \frac{K_i}{s_i^t} - m_i}^{\infty} \frac{1}{\sqrt{2\pi a_i}} e^{-\frac{(x_i)^2}{2a_i}} \, dx_i' \]

\[ = K_i \left( 1 - \Phi\left( \frac{\ln \frac{K_i}{s_i^t} - m_i}{\sqrt{a_i}} \right) \right) \]

\[ = K_i Q\left( \frac{\ln \frac{K_i}{s_i^t} - m_i}{\sqrt{a_i}} \right). \]

Therefore 

\[ u_{a_i}(s_i^t, t) = s_i^t e^{m_i + \frac{a_i}{2}} Q\left( \frac{\ln \frac{K_i}{s_i^t} - (m_i + a_i)}{\sqrt{a_i}} \right) - K_i Q\left( \frac{\ln \frac{K_i}{s_i^t} - m_i}{\sqrt{a_i}} \right). \]

In addition, note that 

\[ u_0(s_i^t, t) = \max(s_i^t e^{m_i} - K_i, 0). \]

We want to find an upper bound of 

\[ |u_{a_i}(s_i^t, t) - u_0(s_i^t, t)| \]

as \( a_i \to 0 \).

We must consider two cases:

1. \( m_i \geq \ln(\frac{K_i}{s_i^t}) \)
Remark. Note that the last inequality uses Proposition 21 (Chernoff bound of the $Q$ function).

As a consequence, there exist constants $C_1^i, C_2^i, C_3^i, C_4^i > 0$ such that

$$|u_{a_i}(s^i_t, t) - u_0(s^i_t, t)| \leq C_i^1(e^{C_2^i a_i} - 1) + C_4^i e^{-C_1^i a_i}.$$  

Case 2: If $m_i < \ln(K_i/s_i)$ then max$(s^i_t e^{m_i} - K_i, 0) = s^i_t e^{m_i} - K_i$. Consequently,

$$|u_{a_i}(s^i_t, t) - u_0(s^i_t, t)| = \left| s^i_t e^{m_i} e^{a_i} (1 - \Phi(\frac{\ln K_i - (m_i + a_i)}{\sqrt{a_i}})) - K_i (1 - \Phi(\frac{\ln K_i - m_i}{\sqrt{a_i}})) \right|$$

$$\leq s^i_t e^{m_i} e^{a_i} (1 - \Phi(\frac{\ln K_i - (m_i + a_i)}{\sqrt{a_i}})) + K_i (1 - \Phi(\frac{\ln K_i - m_i}{\sqrt{a_i}})).$$

for $a_i > 0$ small enough $\ln(K_i/s_i) - (m_i + a_i) > 0$, and

$$\ln \frac{K_i}{s_i} - (m_i + a_i) \sim O \left( \frac{\ln K_i - m_i}{\sqrt{a_i}} \right), \quad \text{as } a_i \to 0.$$

Therefore, using Proposition 21 (Chernoff bound of the $Q$ function) we have that

$$|u_{a_i}(s^i_t, t) - u_0(s^i_t, t)| = \left| s^i_t e^{m_i} e^{a_i} Q(\frac{\ln K_i - (m_i + a_i)}{\sqrt{a_i}}) \right| + K_i Q(\frac{\ln K_i - m_i}{\sqrt{a_i}})$$

$$\leq s^i_t e^{m_i} e^{a_i} e^{-\frac{(\ln K_i - (m_i + a_i))^2}{2 \sigma_i^2}} + K_i e^{-\frac{(\ln K_i - m_i)^2}{2 \sigma_i^2}}$$

$$\leq (s^i_t e^{m_i} e^{a_i} + K_i)e^{-\frac{(\ln K_i - m_i)^2}{2 \sigma_i^2}}$$

$$\sim O(e^{-\frac{N}{a_i}}),$$
with $C > 0$.

Therefore, if $a_j > 0$ is small enough for $j = 1, \ldots, n - k$, then there exist constants $C_1^j, C_2^j, C_3^j, C_4^j > 0$ such that

$$
|u^C_A(s_t, t) - u^C_\tilde{A}(s_t, t)| \leq \sum_{j=1}^{n-k} C_1^j (e^{C_2^j a_{k+j}} - 1) + C_3^j e^{-C_4^j a_{k+j}}.
$$

\[\square\]

**Corollary 25.1.** Let $u^C_A(s_t, t)$ denote the price of a portfolio of $n$ call options (with the same maturity) at time $t$ with positive definite covariance matrix $A$. If the underlying assets are uncorrelated i.e. $A = D_a$ and $a_j$ small enough for $j = k + 1, \cdots, n$, then there exist a constant $C > 0$ such that

$$
|u^C_A(s_t, t) - u^C_{\tilde{A}}(s_t, t)| \leq C \sum_{j=1}^{n-k} a_{k+j},
$$

where $u^C_{\tilde{A}}(s_t, t)$ denotes an approximation of the price obtained by making zero $n-k$ eigenvalues of the covariance matrix i.e $\tilde{A} = D_{\tilde{a}}$ with $\tilde{a} = (a_1, \cdots, a_k, 0, \cdots, 0)$.

Monte Carlo simulations were used to plot the magnitude of the difference between the price of a single asset option and the approximation obtained by making the variance zero. Note that the covariance matrix of a portfolio of single asset options with uncorrelated assets is diagonal, therefore the eigenvalues are the variances of each asset.

Figure 2.2 shows the plot for a digital option. 4000000 simulations were used. The following values were assigned to the parameters and variables of the option: $S_t = 100$, $K = 96$ and $D = 10$, $\Delta = T - t = 1$, $\sigma = 0.2$ and $r = 0$.

**Figure 2.2:** Magnitude of the difference between the price of a digital option and the approximation obtained by making the variance zero.

Figure 2.3 shows the plot for a call option. The following values were assigned to the parameters and variables of the option: $S_t = 100$, $K = 96$, $\Delta = T - t = 1$, $\sigma = 0.2$ and $r = 0.02$. The total number of simulations used was 4000000.
2.4.4 Correlated underlying assets (Two dimensional case)

Let $u_A(s_1^t, s_2^t, t)$ denote the price at time $t$ of an option that depends on two underlying assets with payoff $P(s_1^T, s_2^T)$ and positive definite covariance matrix $A$, then

$$
u_A(s_t, t) = \frac{e^{-r(T-t)}}{2\pi |\text{det}(A)|^{1/2}} \int_{\mathbb{R}^2} P(s_t e^x) \exp\left(-\frac{1}{2}(x - m)'A^{-1}(x - m)\right) dx,$$

where

$$
m = (m_1, m_2),
\quad mi = r - \sigma_i^2 (T - t),
\quad A = (T - t)D\sigma C D\sigma,
\quad s_t = (s_1^t, s_2^t),
\quad s_t e^x = (s_1^t e^{x_1}, s_2^t e^{x_2}).$$

$D\sigma$ is a matrix with $\sigma_i$ along the diagonal and zeros everywhere else, $C$ is a correlation matrix, $C = (\rho_{ij})_{ij}$ and $r$ is the risk free interest rate.

Making the substitution $x = Qy$, where $Q$ is obtained from the eigenvalue decomposition of $A (A = QD\alpha Q')$ we obtain that

$$
u_A(s_t, t) = \frac{e^{-r(T-t)}}{2\pi \prod_{i=1}^2 |a_i|^{1/2}} \int_{\mathbb{R}^2} P(s_t e^{Qy}) \exp\left(-\frac{1}{2}(y - \tilde{m})'D\alpha^{-1}(y - \tilde{m})\right) dy,$$

with $\tilde{m} = Q'm$.

Let

$$Q = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

If $P(S_1^T, S_2^T) = D_1I(S_1^T - K_1 > 0) + D_2I(S_2^T - K_2 > 0)$ (portfolio of two digital options with the same maturity) then

$$
u_A(s_1^t, s_2^t, t) = e^{-r(T-t)} \int_{\mathbb{R}^2} D_1I(s_1^t e^{c_{11}y_1} + c_{12}y_2 - K_1 > 0)p_{a_1, \tilde{m}_1}(y_1)p_{a_2, \tilde{m}_2}(y_2)dy_1dy_2 +
\quad e^{-r(T-t)} \int_{\mathbb{R}^2} D_2I(s_2^t e^{c_{21}y_1} + c_{22}y_2 - K_2 > 0)p_{a_1, \tilde{m}_1}(y_1)p_{a_2, \tilde{m}_2}(y_2)dy_1dy_2.$$
If \( P(S^1_T, S^2_T) = \max(S^1_T - K_1, 0) + \max(S^2_T - K_2, 0) \) (portfolio of two call options with the same maturity) then
\[
u_A(s^1_t, s^2_t, t) = e^{-r(T-t)} \int_{\mathbb{R}^2} \max(s^1_t e^{c_{11} y_1 + c_{12} y_2} - K_1, 0) \rho_{a_1, \tilde{m}_1}(y_1) \rho_{a_2, \tilde{m}_2}(y_2) dy_1 dy_2 + e^{-r(T-t)} \int_{\mathbb{R}^2} \max(s^2_t e^{c_{21} y_1 + c_{22} y_2} - K_2, 0) \rho_{a_1, \tilde{m}_1}(y_1) \rho_{a_2, \tilde{m}_2}(y_2) dy_1 dy_2.
\]

From the eigenvalue decomposition of \( A \), we have that
\[A = QD\alpha Q',\]
where \( \alpha = (a_1, a_2) \) with \( a_i > 0 \) for \( i = 1, 2 \).
Define
\[\tilde{A} = QD\tilde{\alpha} Q',\]
with \( \tilde{\alpha} = (a_1, 0) \).
Let \( u^D_A(s_t, t) \) denote the price at time \( t \) of a portfolio of two digital options then there exists a constant \( C > 0 \) such that
\[|u^D_A(s_t, t) - u^D_A(s_t, t)| \leq C\sqrt{\alpha_2}.
\]
Similarly, Let \( u^C_A(s_t, t) \) denote the price at time \( t \) of a portfolio of two call options then there exist constants \( C_1, C_2, C_3, C_4 > 0 \) such that
\[|u^C_A(s_t, t) - u^C_A(s_t, t)| \leq C_1\sqrt{\alpha_2} + C_2(e^{C_3\alpha_2} - 1) + C_4\alpha_2.
\]
These bounds are formally presented in the next two propositions.

**Proposition 26.** Let \( u^D_A(s_t, t) \) denote the price at time \( t \) of a portfolio of two digital options with the same maturity and correlated underlying assets, with positive definite covariance matrix \( A \). There exists a constant \( C > 0 \) such that
\[|u^D_A(s_t, t) - u^D_A(s_t, t)| \leq C\sqrt{\alpha_2},
\]
where \( u^D_A(s_t, t) \) denotes an approximation of the price obtained by making zero the second eigenvalue of the covariance matrix i.e \( A = QD\alpha Q' \) with \( \alpha = (a_1, a_2) \) and \( \tilde{A} = QD\tilde{\alpha} Q' \) with \( \tilde{\alpha} = (a_1, 0) \).

**Proof.** Let \( u^D_A(s^1_t, s^2_t, t) \) denote the price at time \( t \) of a portfolio of two correlated digital options with positive definite covariance matrix \( A \) and with the same maturity.

Previously, we showed that
\[
u^D_A(s^1_t, s^2_t, t) = e^{-r(T-t)} \int_{\mathbb{R}^2} D_1 \rho(s^1_t e^{c_{11} y_1 + c_{12} y_2} - K_1 > 0) \rho_{a_1, \tilde{m}_1}(y_1) \rho_{a_2, \tilde{m}_2}(y_2) dy_1 dy_2 + e^{-r(T-t)} \int_{\mathbb{R}^2} D_2 \rho(s^2_t e^{c_{21} y_1 + c_{22} y_2} - K_2 > 0) \rho_{a_1, \tilde{m}_1}(y_1) \rho_{a_2, \tilde{m}_2}(y_2) dy_1 dy_2.
\]

Let
\[u^D_A(s^1_t, s^2_t, t) = \lim_{a_2 \to 0} u^D_A(s^1_t, s^2_t, t),\]
and
\[
J_1 = \int_{\mathbb{R}^2} D_1 \rho(s^1_t e^{c_{11} y_1 + c_{12} y_2} - K_1 > 0) \rho_{a_1, \tilde{m}_1}(y_1) \rho_{a_2, \tilde{m}_2}(y_2) dy_1 dy_2
\]
\[
J_2 = \int_{\mathbb{R}^2} D_2 \rho(s^2_t e^{c_{21} y_1 + c_{22} y_2} - K_2 > 0) \rho_{a_1, \tilde{m}_1}(y_1) \rho_{a_2, \tilde{m}_2}(y_2) dy_1 dy_2
\]
\[
J^0_1 = \lim_{a_2 \to 0} J_1
\]
\[
J^0_2 = \lim_{a_2 \to 0} J_2,
\]

then
\[ |u_A^D(s^1_t, s^2_t, t) - u_A^D(s^1_t, s^2_t, t)| \leq e^{-r(T-t)} (|J_1 - J_1^0| + |J_2 - J_2^0|). \]

We just need to find an upper bound of $|J_1 - J_1^0|$ as the same procedure used to bound $|J_1 - J_1^0|$ can be used to obtain an upper bound of $|J_2 - J_2^0|$. Different cases are considered based on the possible values of $c_{1j}$ with $j = 1, 2$.

If $c_{11} \neq 0$ and $c_{12} = 0$ then
\[
J_1 = \int_{\mathbb{R}} D_1 \mathbb{I}(s^1_t e^{c_{11} y_1} - K_1 > 0) p_{a_1, \bar{m}_1}(y_1) p_{a_2, \bar{m}_2}(y_2) dy_1 dy_2 \\
= (\int_{\mathbb{R}} D_1 \mathbb{I}(s^1_t e^{c_{11} y_1} - K_1 > 0) p_{a_1, \bar{m}_1} dy_1) (\int_{\mathbb{R}} p_{a_2, \bar{m}_2}(y_2) dy_2) \\
= \int_{\mathbb{R}} D_1 \mathbb{I}(s^1_t e^{c_{11} y_1} - K_1 > 0) p_{a_1, \bar{m}_1} dy_1.
\]

Therefore
\[
J_1^0 = \int_{\mathbb{R}} D_1 \mathbb{I}(s^1_t e^{c_{11} y_1} - K_1 > 0) p_{a_1, \bar{m}_1} dy_1,
\]
and $|J_1 - J_1^0| = 0$ for all $a_2 > 0$.

If $c_{11} = 0$ and $c_{12} \neq 0$ then
\[
J_1 = \int_{\mathbb{R}} D_1 \mathbb{I}(s^1_t e^{c_{12} y_2} - K_1 > 0) p_{a_1, \bar{m}_1}(y_1) p_{a_2, \bar{m}_2}(y_2) dy_1 dy_2 \\
= (\int_{\mathbb{R}} D_1 \mathbb{I}(s^1_t e^{c_{12} y_2} - K_1 > 0) p_{a_2, \bar{m}_2} dy_2) (\int_{\mathbb{R}} p_{a_1, \bar{m}_1}(y_1) dy_1) \\
= \int_{\mathbb{R}} D_1 \mathbb{I}(s^1_t e^{c_{12} y_2} - K_1 > 0) p_{a_2, \bar{m}_2} dy_2.
\]

Therefore there exist constants $C_1, C_2 > 0$ such that $|J_1 - J_1^0| \leq C_1 e^{-\frac{C_2}{y_2}}$. (This result was proved in the previous section).

We can now consider the case $c_{11}, c_{12} \neq 0$. To fix ideas and illustrate the procedure assume that $c_{11}, c_{12} > 0$, we obtain that
\[
s^1_t e^{c_{11} y_1} + c_{12} y_2 - K_1 > 0 \iff y_2 > \frac{\ln \frac{K_1}{s^1_t}}{\frac{c_{11}}{c_{12}}} - \frac{c_{11}}{c_{12}} y_1.
\]

Therefore
\[
J_1 = \int_{-\infty}^{\infty} \left( \int_{\ln \frac{K_1}{s^1_t}}^{\frac{c_{11}}{c_{12}} - \frac{c_{11}}{c_{12}} y_1} D_1 p_{a_2, \bar{m}_2}(y_2) dy_2 \right) p_{a_1, \bar{m}_1}(y_1) dy_1 \\
= D_1 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi a_1}} e^{-\frac{(y_1 - \bar{m}_1)^2}{2a_1}} \left( \int_{\ln \frac{K_1}{s^1_t}}^{\frac{c_{11}}{c_{12}} - \frac{c_{11}}{c_{12}} y_1} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2 \right) dy_1.
\]

Hence
\[
\lim_{a_2 \to 0} J_1 = D_1 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi a_1}} e^{-\frac{(y_1 - \bar{m}_1)^2}{2a_1}} \left( \lim_{a_2 \to 0} \int_{\ln \frac{K_1}{s^1_t}}^{\frac{c_{11}}{c_{12}} - \frac{c_{11}}{c_{12}} y_1} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2 \right) dy_1.
\]

Note that we can exchange the integrals using Theorem 17 (Tonelli-Fubini) and pass the limit inside the integral using Theorem 15 (Dominated Convergence). We must determine
\[
\lim_{a_2 \to 0} \int_{\ln \frac{K_1}{s^1_t}}^{\frac{c_{11}}{c_{12}} - \frac{c_{11}}{c_{12}} y_1} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2.
\]
Note that
\[
\int_{\ln \frac{K_2}{c_1}}^{\infty} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2.
\]

Therefore
\[
\lim_{\alpha_2 \to 0} \int_{\ln \frac{K_2}{c_1}}^{\infty} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2
\]
is
\[
\begin{cases}
1, & \text{if } \bar{m}_2 > \frac{\ln K_2}{c_2} - \frac{c_{11}}{c_{12}} y_1, \\
\frac{1}{2}, & \text{if } \bar{m}_2 = \frac{\ln K_2}{c_2} - \frac{c_{11}}{c_{12}} y_1, \\
0, & \text{if } \bar{m}_2 < \frac{\ln K_2}{c_2} - \frac{c_{11}}{c_{12}} y_1.
\end{cases}
\]

Note that as we are assuming that \(c_{11}, c_{12} > 0\) then
\[
\bar{m}_2 > \frac{\ln K_2}{c_2} - \frac{c_{11}}{c_{12}} y_1 \iff y_1 > \frac{\ln K_2}{c_1} - \frac{c_{12}}{c_{11}} \bar{m}_2,
\]

As a consequence,
\[
J_1^0 = \lim_{a_2 \to 0} J_1 = D_1 \int_{\ln \frac{K_2}{c_1}}^{\infty} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2.
\]

In addition,
\[
J_2 = I_1 + I_2,
\]
where
\[
I_1 = D_1 \int_{-\infty}^{\ln \frac{K_2}{c_1}} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2 p_{a_2, \bar{m}_2}(y_2) dy_2.
\]

and
\[
I_2 = D_1 \int_{\ln \frac{K_2}{c_1}}^{\infty} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2 p_{a_2, \bar{m}_2}(y_2) dy_2.
\]

Let \(I_1^0 = \lim_{a_2 \to 0} I_1\) and \(I_2^0 = \lim_{a_2 \to 0} I_2\) then
\[
|J_1 - J_1^0| \leq |I_1 - I_1^0| + |I_2 - I_2^0|,
\]

with \(I_1^0 = 0\) and \(I_2^0 = D_1 \int_{\ln \frac{K_2}{c_1}}^{\infty} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2.
\]

Note that
\[
I_1 = D_1 \int_{-\infty}^{\ln \frac{K_2}{c_1}} \frac{1}{\sqrt{2\pi a_2}} e^{-\frac{(y_2 - \bar{m}_2)^2}{2a_2}} dy_2 p_{a_2, \bar{m}_2}(y_2) dy_2,
\]
with
\[
\frac{\ln K_2}{c_2} - \bar{m}_2 = \frac{c_{11}}{c_{12}} y_1 > 0.
\]
for $y_1 \leq \frac{\ln K_1}{c_{11}} - \frac{c_{12}}{c_{11}} \bar{m}_2$.

In addition,

$$|I_2 - I_2^0| = D_1 \left| \int_{\ln K_1 \over c_{11}}^{\infty} p_{a_2, \bar{m}_2} (y_2) dy_2 \right| p_{a_1, \bar{m}_1} (y_1) dy_1$$

$$= D_1 \int_{\ln K_1 \over c_{11}}^{\infty} 1 - \frac{c_{12}}{c_{11}} \bar{m}_2 \left( \frac{c_{11}}{c_{12}} y_1 - \frac{c_{11}}{c_{12}} \bar{m}_2 \right) \sqrt{\alpha_2} p_{a_1, \bar{m}_1} (y_1) dy_1$$

$$= D_1 \int_{\ln K_1 \over c_{11}}^{\infty} Q \left( \frac{c_{11}}{c_{12}} y_1 - \frac{c_{11}}{c_{12}} \bar{m}_2 \right) \sqrt{\alpha_2} p_{a_1, \bar{m}_1} (y_1) dy_1.$$

So far we have shown that under the assumption of $c_{11}, c_{12} > 0$,

$$|J_1 - J_1^0| \leq T_1 + T_2$$

where

$$T_1 = D_1 \int_{-\infty}^{\ln K_1 \over c_{11}} - \frac{c_{12}}{c_{11}} \bar{m}_2 Q \left( \frac{\ln K_1 \over c_{11}}{c_{12}} - \bar{m}_2 - \frac{c_{11}}{c_{12}} y_1 \right) \sqrt{\alpha_2} p_{a_1, \bar{m}_1} (y_1) dy_1$$

$$T_2 = D_1 \int_{\ln K_1 \over c_{11}}^{\infty} Q \left( \frac{c_{11}}{c_{12}} y_1 - \frac{c_{11}}{c_{12}} \bar{m}_2 \right) \sqrt{\alpha_2} p_{a_1, \bar{m}_1} (y_1) dy_1.$$

On the other hand, define

$$T_3 = D_1 \int_{-\infty}^{\ln K_1 \over c_{11}} - \frac{c_{12}}{c_{11}} \bar{m}_2 Q \left( \frac{\ln K_1 \over c_{11}}{c_{12}} - \bar{m}_2 - \frac{c_{11}}{c_{12}} y_1 \right) \sqrt{\alpha_2} p_{a_1, \bar{m}_1} (y_1) dy_1$$

and

$$T_4 = D_1 \int_{\ln K_1 \over c_{11}}^{\infty} Q \left( \frac{\ln K_1 \over c_{11}}{c_{12}} - \bar{m}_2 - \frac{c_{11}}{c_{12}} y_1 \right) \sqrt{\alpha_2} p_{a_1, \bar{m}_1} (y_1) dy_1.$$

In general, for the remaining cases:

- $c_{11} > 0$ and $c_{12} < 0$
- $c_{11} < 0$ and $c_{12} > 0$
- $c_{11} < 0$ and $c_{12} < 0$

we obtain that

$$|J_1 - J_1^0| \leq \begin{cases} T_1 + T_2, & \text{if } c_{11} \over c_{12} > 0, \\ T_3 + T_4, & \text{if } c_{11} \over c_{12} < 0. \end{cases}$$
The proofs of the remaining cases are almost identical to the proof provided above of the case: \( c_{11} > 0 \) and \( c_{12} > 0 \). In addition, we have that:

\[
T_i \leq \begin{cases}
D_1 \int_{-\infty}^{\infty} \frac{\ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2}{\sqrt{2\pi a_1}} \frac{1}{e^{-\frac{\ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2}{2a_2}} e^{-\frac{(y_1 - \tilde{m}_1)^2}{2a_1}} dy_1, & \text{for } i = 1, 3.
\end{cases}
\]

\[
D_1 \int_{\ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2}^{\infty} \frac{1}{\sqrt{2\pi a_1}} \frac{1}{e^{-\frac{\ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2}{2a_2}} e^{-\frac{(y_1 - \tilde{m}_1)^2}{2a_1}} dy_1, \quad \text{for } i = 2, 4.
\]

using Proposition 21 (Chernoff bound of the \( Q \) function).

Next, we obtain an upper bound for the first integral:

\[
R = D_1 \int_{-\infty}^{\ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2} \frac{1}{\sqrt{2\pi a_1}} e^{-\frac{\ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2}{2a_2}} e^{-\frac{(y_1 - \tilde{m}_1)^2}{2a_1}} dy_1.
\]

In particular, we show that there exists a constant \( C > 0 \) such that

\[
D_1 \int_{-\infty}^{\ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2} \frac{1}{\sqrt{2\pi a_1}} e^{-\frac{\ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2}{2a_2}} e^{-\frac{(y_1 - \tilde{m}_1)^2}{2a_1}} dy_1 \leq C \sqrt{a_2}
\]

for \( a_2 > 0 \) small enough.

Note that there also exists a constant \( C' \) such that for the second integral

\[
D_1 \int_{\ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2}^{\infty} \frac{1}{\sqrt{2\pi a_1}} e^{-\frac{\ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2}{2a_2}} e^{-\frac{(y_1 - \tilde{m}_1)^2}{2a_1}} dy_1 \leq C' \sqrt{a_2}
\]

for \( a_2 > 0 \) small enough. The proof of this result is almost identical to the proof for the first integral shown below.

Let

\[
A = \ln \frac{K_i}{\sqrt{c_{11}}} - \frac{c_{12}}{c_{11}} \tilde{m}_2
\]

\[
d = \ln \frac{K_i}{\sqrt{c_{11}}} - \tilde{m}_2
\]

\[
c = \frac{c_{11}}{c_{11}} - \tilde{m}_2
\]

\[
b = 2a_2
\]

\[
a = 2a_1
\]

\[
m = \tilde{m}_1.
\]

From Proposition 23,

\[
\int_{-\infty}^{A} e^{-\frac{(d-cy_1)^2}{2}} e^{-\frac{(y_1-m)^2}{2}} dy_1 = e^{-\frac{(c_1-\hat{m}_1)^2}{2a_2}} \int_{-\infty}^{A} e^{-\frac{(x+2a_1)^2}{2a_2}} (y_1 - \frac{a_1+bx_1}{a_2+a_1+b})^2 dy_1.
\]
Therefore there exists a constant $C > 0$ such that
\[
R \leq C \int_{-\infty}^{A} \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2 + b}{2}(y_1 - \frac{acd + bm}{ac^2 + b})^2} dy_1.
\]

Consider the integral
\[
\int_{-\infty}^{A} \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2 + b}{2}(y_1 - \frac{acd + bm}{ac^2 + b})^2} dy_1.
\]

Let
\[
\tilde{y}_1 = \frac{\sqrt{ac^2 + b}}{\sqrt{ab}} \left( y_1 - \frac{acd + bm}{ac^2 + b} \right),
\]
then $dy_1 = \frac{\sqrt{ab}}{\sqrt{2\sqrt{ac^2 + b}}} d\tilde{y}_1$ and at $y_1 = A$, $\tilde{y}_1 = \tilde{A} = \frac{\sqrt{2\sqrt{ac^2 + b}}}{\sqrt{ab}} (A - \frac{acd + bm}{ac^2 + b})$.

Therefore,
\[
\int_{-\infty}^{A} \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2 + b}{2}(y_1 - \frac{acd + bm}{ac^2 + b})^2} dy_1 = \frac{\sqrt{ab}}{\sqrt{2\sqrt{ac^2 + b}}} \int_{-\infty}^{\tilde{A}} e^{-\frac{\tilde{y}_1^2}{2}} d\tilde{y}_1 = \frac{\sqrt{ab}}{\sqrt{2\sqrt{ac^2 + b}}} \Phi(\tilde{A}),
\]
where $\tilde{A} = \frac{\sqrt{2\sqrt{ac^2 + b}}}{\sqrt{ab}} (A - \frac{acd + bm}{ac^2 + b})$.

Returning to our original variables we obtain that
\[
\tilde{A} = \frac{\sqrt{2\sqrt{ac^2 + b}}}{\sqrt{2a_1}} \left[ \left( \frac{\ln k_1}{c_{11}} - \frac{c_{12}}{c_{11}} \tilde{m}_2 \right) - \frac{2a_1 \frac{c_{11}}{c_{12}} - \frac{c_{12}}{c_{11}} \tilde{m}_2}{2a_1 \frac{c_{11}}{c_{12}} + 2a_2} \right]
\]
\[
= \frac{\sqrt{2\sqrt{ac^2 + b}}}{\sqrt{2a_1}} \left( \frac{\ln k_1}{c_{11}} - \frac{c_{12}}{c_{11}} \tilde{m}_2 - \tilde{m}_1 \right),
\]
\[
= \frac{\sqrt{2\sqrt{ac^2 + b}}}{\sqrt{2a_1}} \left( \frac{\ln k_1}{c_{11}} - \frac{c_{12}}{c_{11}} \tilde{m}_2 - \tilde{m}_1 \right),
\]
and
\[
\frac{\sqrt{b}}{\sqrt{ac^2 + b}} = \frac{\sqrt{a_2}}{\sqrt{a_1 (\frac{c_{11}}{c_{12}})^2 + a_2}}.
\]

As a consequence, as $a_2 \to 0$
\[
\frac{\sqrt{a_2}}{\sqrt{a_1 (\frac{c_{11}}{c_{12}})^2 + a_2}} \Phi(\tilde{A}) \sim O(a_2^{\frac{1}{2}}).
\]

Hence, there exits $C > 0$ such that
\[
R \leq C \sqrt{a_2}.
\]

**Proposition 27.** Let $u_A^C(s_t, t)$ denote the price at time $t$ of a portfolio of two call options with the same maturity and correlated underlying assets with positive definite covariance matrix $A$. There exist constants $C_1, C_2, C_3, C_4 > 0$ such that
\[
|u_A^C(s_t, t) - u_A(s_t, t)| \leq C_1 \sqrt{a_2} + C_2 (e^{C_3 a_2} - 1) + C_4 a_2,
\]
where $u_A(s_t, t)$ denotes an approximation of the price obtained by making zero the second eigenvalue of the
covariance matrix i.e. $A = QD_A Q'$ with $a = (a_1, a_2)$ and $\tilde{A} = QD_\tilde{a} Q'$ with $\tilde{a} = (a_1, 0)$.

**Proof.** Let $u^C_A(s^1_t, s^2_t, t)$ denote the price at time $t$ of a portfolio of two correlated call options with positive definite covariance matrix $A$ and with the same maturity.

Previously, we showed that

$$ u^C_A(s^1_t, s^2_t, t) = e^{-r(T-t)} \int_{\mathbb{R}^2} \max(s_1^1 e^{c_1 y_1 + c_2 y_2} - K_1, 0) p_{a_1, \tilde{m}_1} (y_1) p_{a_2, \tilde{m}_2} (y_2) dy_1 dy_2 + e^{-r(T-t)} \int_{\mathbb{R}^2} \max(s_1^2 e^{c_1 y_1 + c_2 y_2} - K_2, 0) p_{a_1, \tilde{m}_1} (y_1) p_{a_2, \tilde{m}_2} (y_2) dy_1 dy_2. $$

Let

$$ u^C_A(s^1_t, s^2_t, t) = \lim_{a_2 \to 0} u^C_A(s^1_t, s^2_t, t), $$

and

$$ J_1 = \int_{\mathbb{R}^2} \max(s_1^1 e^{c_1 y_1 + c_2 y_2} - K_1, 0) p_{a_1, \tilde{m}_1} (y_1) p_{a_2, \tilde{m}_2} (y_2) dy_1 dy_2 $$

$$ J_2 = \int_{\mathbb{R}^2} \max(s_1^2 e^{c_1 y_1 + c_2 y_2} - K_2, 0) p_{a_1, \tilde{m}_1} (y_1) p_{a_2, \tilde{m}_2} (y_2) dy_1 dy_2 $$

$$ J^0_1 = \lim_{a_2 \to 0} J_1 $$

$$ J^0_2 = \lim_{a_2 \to 0} J_2. $$

then

$$ |u^C_A(s^1_t, s^2_t, t) - u^C_A(s^1_t, s^2_t, t)| \leq e^{-r(T-t)} (|J_1 - J^0_1| + |J_2 - J^0_2|). $$

We just need to find an upper bound of $|J_1 - J^0_1|$, as the same procedure used to bound $|J_1 - J^0_1|$ can be used to obtain an upper bound of $|J_2 - J^0_2|$.

Different cases are considered based on the possible values of $c_{1j}$ with $j = 1, 2$.

If $c_{11} \neq 0$ and $c_{12} = 0$ then

$$ J_1 = \int_{\mathbb{R}^2} \max(s_1^1 e^{c_1 y_1} - K_1, 0) p_{a_1, \tilde{m}_1} (y_1) p_{a_2, \tilde{m}_2} (y_2) dy_1 dy_2 $$

$$ = \left( \int_{\mathbb{R}} \max(s_1^1 e^{c_1 y_1} - K_1, 0) p_{a_1, \tilde{m}_1} (y_1) dy_1 \right) \left( \int_{\mathbb{R}} p_{a_2, \tilde{m}_2} (y_2) dy_2 \right) $$

$$ = \int_{\mathbb{R}} \max(s_1^1 e^{c_1 y_1} - K_1, 0) p_{a_1, \tilde{m}_1} (y_1) dy_1. $$

Therefore

$$ J^0_1 = \int_{\mathbb{R}} \max(s_1^1 e^{c_1 y_1} - K_1, 0) p_{a_1, \tilde{m}_1} (y_1), $$

and $|J_1 - J^0_1| = 0$ for all $a_2 > 0$.

If $c_{11} = 0$ and $c_{12} \neq 0$ then

$$ J_1 = \int_{\mathbb{R}^2} \max(s_1^1 e^{c_2 y_2} - K_1, 0) p_{a_1, \tilde{m}_1} (y_1) p_{a_2, \tilde{m}_2} (y_2) dy_1 dy_2 $$

$$ = \left( \int_{\mathbb{R}} \max(s_1^1 e^{c_2 y_2} - K_1, 0) p_{a_2, \tilde{m}_2} (y_2) dy_2 \right) \left( \int_{\mathbb{R}} p_{a_1, \tilde{m}_1} (y_1) dy_1 \right) $$

$$ = \int_{\mathbb{R}} \max(s_1^1 e^{c_2 y_2} - K_1, 0) p_{a_2, \tilde{m}_2} (y_2) dy_2. $$

Therefore there exist constants $C_1, C_2, C_3, C_4 > 0$ such that

$$ |J_1 - J^0_1| \leq C_1 (e^{C_2 a_2} - 1) + C_3 e^{-C_4 a_2}. $$

This result was proved in the previous section (portfolio of uncorrelated call options).

We can now consider the case $c_{11}, c_{12} \neq 0$. To fix ideas and illustrate the procedure assume that $c_{11}, c_{12} > 0$, we obtain that

$$ s_1^1 e^{c_1 y_1 + c_2 y_2} - K_1 > 0 \iff y_2 > \frac{\ln \frac{K_1}{s_1^1}}{c_2} - \frac{c_{11}}{c_{12}} y_1. $$
Therefore,

\[
J_1 = \int_{-\infty}^{\infty} \left( \int_{\ln K_1 \over c_{11}}^{\infty} \frac{e^{c_{11} y_1 + c_{12} y_2} - K_1}{c_{11} y_1} \right) p_{a_1, \tilde{m}_1}(y_1) dy_1 \text{ for } y_2 \geq \frac{\ln \frac{c_{11}}{K_1}}{c_{12}} - \frac{c_{11}}{c_{12}} y_1,
\]

\[
\text{and as a consequence of Theorem 15 (Dominated Convergence)} \quad J_1^0 = \lim_{a_2 \to 0} J_1
\]

\[
= \int_{-\infty}^{\infty} p_{a_1, \tilde{m}_1}(y_1) \left( \lim_{a_2 \to 0} \int_{\ln K_1 \over c_{11}}^{\infty} \frac{e^{c_{11} y_1 + c_{12} y_2} - K_1}{c_{11} y_1} \right) dy_1.
\]

However,

\[
\lim_{a_2 \to 0} \int_{\ln K_1 \over c_{11}}^{\infty} \frac{1}{2\pi a_2} \left( \frac{e^{c_{11} y_1 + c_{12} y_2} - K_1 e^{-\frac{(y_2 - \tilde{m}_2)^2}{2a_2^2}}}{c_{11} y_1} \right) dy_1
\]

is

\[
\begin{cases}
  s_1 e^{c_{11} y_1 + c_{12} \tilde{m}_2} - K_1, & \text{if } \tilde{m}_2 \geq \frac{\ln \frac{c_{11}}{K_1}}{c_{12}} - \frac{c_{11}}{c_{12}} y_1, \\
  0, & \text{if } \tilde{m}_2 < \frac{\ln \frac{c_{11}}{K_1}}{c_{12}} - \frac{c_{11}}{c_{12}} y_1.
\end{cases}
\]

Therefore

\[
J_1^0 = \int_{-\infty}^{\infty} \frac{e^{c_{11} y_1 + c_{12} \tilde{m}_2}}{c_{11} \tilde{m}_2} (s_1 e^{c_{11} y_1 + c_{12} y_2} - K_1) p_{a_1, \tilde{m}_1}(y_1) dy_1.
\]

In addition, \( J_1 = I_1 + I_2 \) with

\[
I_1 = \int_{-\infty}^{\infty} \frac{e^{c_{11} y_1 + c_{12} \tilde{m}_2}}{c_{11} \tilde{m}_2} \left( \int_{\ln K_1 \over c_{11}}^{\infty} \frac{e^{c_{11} y_1 + c_{12} y_2} - K_1}{c_{11} y_1} \right) p_{a_1, \tilde{m}_1}(y_1) dy_1
\]

and

\[
I_2 = \int_{\ln K_1 \over c_{11}}^{\infty} \frac{e^{c_{11} y_1 + c_{12} \tilde{m}_2}}{c_{11} \tilde{m}_2} \left( \int_{\ln K_1 \over c_{11}}^{\infty} \frac{e^{c_{11} y_1 + c_{12} y_2} - K_1}{c_{11} y_1} \right) p_{a_1, \tilde{m}_1}(y_1) dy_1.
\]

Let \( I_1^0 = \lim_{a_2 \to 0} I_1 \) and \( I_2^0 = \lim_{a_2 \to 0} I_2 \) then

\[
J_1^0 = 0
\]

\[
J_2^0 = \int_{\ln K_1 \over c_{11}}^{\infty} \frac{e^{c_{11} y_1 + c_{12} \tilde{m}_2}}{c_{11} \tilde{m}_2} (s_1 e^{c_{11} y_1 + c_{12} y_2} - K_1) p_{a_1, \tilde{m}_1}(y_1) dy_1,
\]

and

\[
|J_1 - J_1^0| \leq |I_1 - I_1^0| + |I_2 - I_2^0|.
\]
Therefore,

\[ |I_1 - I_1^0| = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( s_1^t e^{(c_{11}y_1 + c_{12}y_2)} - K_1 \right) p_{a_2, \tilde{m}_2}(y_2) dy_2 \left| p_{a_1, \tilde{m}_1}(y_1) \right| dy_1. \]

Hence

\[ |I_1 - I_1^0| \leq T_1 + T_2, \]

where

\[ T_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( s_1^t e^{(c_{11}y_1 + c_{12}y_2)} - K_1 \right) p_{a_2, \tilde{m}_2}(y_2) dy_2 \left| p_{a_1, \tilde{m}_1}(y_1) \right| dy_1, \]

and

\[ T_2 = K_1 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( s_1^t e^{(c_{11}y_1 + c_{12}y_2)} - K_1 \right) p_{a_2, \tilde{m}_2}(y_2) dy_2 \left| p_{a_1, \tilde{m}_1}(y_1) \right| dy_1. \]

However, from Proposition 22

\[ \int_{t}^{\infty} e^{c_{12}y_2} p_{a_2, \tilde{m}_2}(y_2) dy_2 = e^{c_{12}y_2} e^{(c_{12})^2 a_2} \int_{t}^{\infty} \frac{1}{\sqrt{2\pi} a_2} \exp \left( \frac{-(y_2 - (\tilde{m}_2 + c_{12}a_2))^2}{2a_2^2} \right) dy_2. \]

Therefore,

\[ T_1 = s_1^t e^{c_{12}y_2} e^{(c_{12})^2 a_2} \int_{-\infty}^{\infty} e^{c_{11}y_1} Q \left( \frac{\left( \ln \frac{K_1}{\xi_{12}} - (\tilde{m}_2 + c_{12}a_2) - c_{11}y_1 \right)}{\sqrt{a_2}} \right) p_{a_1, \tilde{m}_1}(y_1) dy_1, \]

and

\[ T_2 = K_1 \int_{-\infty}^{\infty} e^{c_{12}y_2} e^{(c_{12})^2 a_2} Q \left( \frac{\left( \ln \frac{K_1}{\xi_{12}} - \tilde{m}_2 - c_{11}y_1 \right)}{\sqrt{a_2}} \right) p_{a_1, \tilde{m}_1}(y_1) dy_1. \]

Similarly,

\[ I_2 - I_2^0 = \int_{t}^{\infty} e^{c_{12}y_2} Q \left( \frac{\left( \ln \frac{K_1}{\xi_{12}} - (\tilde{m}_2 + c_{12}a_2) - c_{11}y_1 \right)}{\sqrt{a_2}} \right) p_{a_1, \tilde{m}_1}(y_1) dy_1. \]

Therefore

\[ |I_2 - I_2^0| \leq R_1 + R_2, \]
where

\[ R_1 = \left| \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^1 y y_{11} + c_{12} \tilde{m}_2}{c_{11} - c_{12} \tilde{m}_2} \left( s^1 y y_{11} + c_{12} \tilde{m}_2 - \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^2 e_{12} (y_2)}{c_{12} y_1} p_{a_2, \tilde{m}_2} (y_2) dy_2 \right) p_{a_1, \tilde{m}_1} (y_1) dy_1 \right| \]

and

\[ R_2 = K_1 \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^1 y y_{11} + c_{12} \tilde{m}_2}{c_{11} - c_{12} \tilde{m}_2} \left( 1 - \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^2 e_{12} (y_2)}{c_{12} y_1} p_{a_2, \tilde{m}_2} (y_2) dy_2 \right) p_{a_1, \tilde{m}_1} (y_1) dy_1. \]

Note that

\[ R_1 = \left| \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^1 y y_{11} + c_{12} \tilde{m}_2}{c_{11} - c_{12} \tilde{m}_2} \left( s^1 y y_{11} + c_{12} \tilde{m}_2 - \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^2 e_{12} (y_2)}{c_{12} y_1} p_{a_2, \tilde{m}_2} (y_2) dy_2 \right) p_{a_1, \tilde{m}_1} (y_1) dy_1 \right|. \]

However, using Proposition 22 and the definition of the Q integral,

\[ e^{c_{12} \tilde{m}_2} - \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^1 y y_{11} + c_{12} \tilde{m}_2}{c_{11} - c_{12} \tilde{m}_2} e^{c_{12} y_{22}} p_{a_2, \tilde{m}_2} (y_2) dy_2 \]

can be rewritten as

\[ e^{c_{12} \tilde{m}_2} \left( 1 - e^{\frac{(c_{12})^2 y_{22}}{2}} Q \left( \frac{\ln \frac{K_1}{c_{11}} - (\tilde{m}_2 + c_{12} a_2)}{\sqrt{a_2}} \right) \right). \]

Therefore, as \( Q(x) = 1 - Q(-x) \), we obtain that

\[ e^{c_{12} \tilde{m}_2} - \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^1 y y_{11} + c_{12} \tilde{m}_2}{c_{11} - c_{12} \tilde{m}_2} e^{c_{12} y_{22}} p_{a_2, \tilde{m}_2} (y_2) dy_2 = e^{c_{12} \tilde{m}_2} \left( 1 - e^{\frac{(c_{12})^2 y_{22}}{2}} \right) + e^{c_{12} \tilde{m}_2} e^{\frac{(c_{12})^2 y_{22}}{2}} Q \left( \frac{\ln \frac{K_1}{c_{11}} - (\tilde{m}_2 + c_{12} a_2)}{\sqrt{a_2}} \right). \]

Hence,

\[ R_1 \leq e^{c_{12} \tilde{m}_2} \left| 1 - e^{\frac{(c_{12})^2 y_{22}}{2}} \right| \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^1 y y_{11} + c_{12} \tilde{m}_2}{c_{11} - c_{12} \tilde{m}_2} \left( s^1 y y_{11} + c_{12} \tilde{m}_2 - \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^2 e_{12} (y_2)}{c_{12} y_1} p_{a_2, \tilde{m}_2} (y_2) dy_2 \right) p_{a_1, \tilde{m}_1} (y_1) dy_1 + \]

\[ s^1 e^{c_{12} \tilde{m}_2} e^{\frac{(c_{12})^2 y_{22}}{2}} \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^1 y y_{11} + c_{12} \tilde{m}_2}{c_{11} - c_{12} \tilde{m}_2} e^{c_{11} y_{11} Q} \left( \frac{\ln \frac{K_1}{c_{11}} - (\tilde{m}_2 + c_{12} a_2)}{\sqrt{a_2}} \right) p_{a_1, \tilde{m}_1} (y_1) dy_1. \]

In addition, note that

\[ R_2 = K_1 \int_{\ln \frac{K_1}{c_{11}}}^{\infty} \frac{s^1 y y_{11} + c_{12} \tilde{m}_2}{c_{11} - c_{12} \tilde{m}_2} Q \left( \frac{\ln \frac{K_1}{c_{11}} - (\tilde{m}_2 + c_{12} a_2)}{\sqrt{a_2}} \right) p_{a_1, \tilde{m}_1} (y_1) dy_1. \]
To sum up, we have that

$$ |J_1 - J_1^0| \leq \sum_{i=1}^{5} T_i, $$

where

$$ T_1 = s_1^1 e^{c_{12} \tilde{m}_2} e^{\frac{(c_{12})^2 a_2}{2}} \int_{-\infty}^{\infty} e^{c_{11} y_1 Q} \left( \frac{\left( \frac{\ln K_1}{\sqrt{c_{12}}} - \tilde{m}_2 \right)^2}{\sqrt{\sigma_2}} \right) p_{a_1, \tilde{m}_1} (y_1) dy_1 $$

$$ T_2 = K_1 \int_{-\infty}^{\infty} e^{c_{12} \tilde{m}_2} \left( \frac{\left( \frac{\ln K_1}{\sqrt{c_{12}}} - \tilde{m}_2 \right)}{\sqrt{\sigma_2}} \right) p_{a_1, \tilde{m}_1} (y_1) dy_1 $$

$$ T_3 = e^{c_{12} \tilde{m}_2} \left( 1 - e^{(c_{12})^2 a_2} \right) \int_{\ln K_1}^{\infty} e^{c_{11} y_1} p_{a_1, \tilde{m}_1} (y_1) dy_1 $$

$$ T_4 = s_1^1 e^{c_{12} \tilde{m}_2} e^{\frac{(c_{12})^2 a_2}{2}} \int_{\ln K_1}^{\infty} e^{c_{11} y_1 Q} \left( \frac{\left( \frac{\ln K_1}{\sqrt{c_{12}}} - \tilde{m}_2 \right)^2}{\sqrt{\sigma_2}} \right) p_{a_1, \tilde{m}_1} (y_1) dy_1 $$

and

$$ T_5 = K_1 \int_{-\infty}^{\infty} e^{c_{11} y_1 Q} \left( \frac{c_{11} y_1 - \left( \frac{\ln K_1}{\sqrt{c_{12}}} - \tilde{m}_2 \right)}{\sqrt{\sigma_2}} \right) p_{a_1, \tilde{m}_1} (y_1) dy_1. $$

Note that previously (portfolio of two Digital options) we showed that there exists a constant $C_1 > 0$ such that $T_i \leq C_1 \sqrt{a_2}$ for $a_2 > 0$ small enough and $i = 2, 5$.

In addition, there exists constants $C_2, C_3 > 0$ such that $T_3 \leq C_2 (e^{C_3 a_2} - 1)$. Therefore, we just need to find upper bounds of $T_1$ and $T_4$. We only show the proof for $T_3$ as the proof for $T_4$ is almost identical.

Consider the following integral

$$ S = \int_{-\infty}^{\infty} e^{c_{11} y_1 Q} \left( \frac{\left( \frac{\ln K_1}{\sqrt{c_{12}}} - \tilde{m}_2 \right)^2}{\sqrt{\sigma_2}} \right) p_{a_1, \tilde{m}_1} (y_1) dy_1. $$

From Proposition 22, we have that

$$ S = C \int_{-\infty}^{\infty} e^{c_{11} y_1 Q} \left( \frac{\left( \frac{\ln K_1}{\sqrt{c_{12}}} - \tilde{m}_2 \right)^2}{\sqrt{\sigma_2}} \right) dy_1 $$

with $C = e^{\tilde{m}_1 c_{11} + \frac{(c_{11})^2}{2} a_1}$.

Splitting the integration interval we obtain that

$$ S = CS_1 + CS_2, $$
Therefore, 
\[ S_1 = \int_{-\infty}^{\infty} e^{-\frac{(y - \tilde{m}_1 - c_{11}a_2)^2}{2a_1^2}} \left( \frac{1}{\sqrt{2\pi a_1}} \right) dy_1 \]
\[ S_2 = \int_{\infty}^{\infty} e^{-\frac{(y - (\tilde{m}_1 + c_{11}a_1))^2}{2a_1^2}} \left( \frac{1}{\sqrt{2\pi a_1}} \right) dy_1. \]

Note that the integrand in \( S_2 \) is bounded by 1. Therefore, there exists a constant \( C_4 > 0 \) such that \( S_2 \leq C_4 a_2 \).

Using Proposition 21 (Chernoff bound of the \( Q \) function), we obtain that
\[ S_1 \leq \int_{-\infty}^{\infty} e^{-\frac{(\tilde{m}_1 - \tilde{m}_2 - c_{12}a_2)^2}{2a_2^2}} \left( \frac{1}{\sqrt{2\pi a_2}} \right) dy_1. \]

Define:
\[ A = \frac{\ln K_1}{c_1} - \frac{c_{12} \tilde{m}_2}{c_{11}} - \frac{c_{12}}{c_{11}} a_2 \]
\[ d = \frac{\ln K_1}{c_{11}} - \tilde{m}_2 - c_{12}a_2 \]
\[ c = \frac{c_{11}}{c_{12}} \]
\[ b = 2a_2 \]
\[ a = 2a_1 \]
\[ m = \tilde{m}_1 + c_{11}a_1. \]

Using Proposition 23,
\[ \int_{-\infty}^{A} e^{-\frac{(x-c_0)^2}{a^2}} e^{-\frac{(y-m)^2}{b^2}} dy_1 = e^{-\frac{(c_0 - m)^2}{b^2}} \int_{-\infty}^{A} e^{-\frac{a^2(x-c_0)^2}{a^2}} dy_1. \]

Therefore, there exists a constant \( C > 0 \) such that
\[ S_1 \leq C \int_{-\infty}^{A} \frac{1}{\sqrt{a\pi}} e^{-\frac{a^2}{a^2} (y_1 - \frac{a cd + bm}{ac^2 + b})^2} dy_1. \]

Consider the integral
\[ \int_{-\infty}^{A} \frac{1}{\sqrt{a\pi}} e^{-\frac{a^2}{a^2} (y_1 - \frac{a cd + bm}{ac^2 + b})^2} dy_1. \]

Let
\[ \frac{\tilde{y}_1}{\sqrt{2}} = \frac{\sqrt{ac^2 + b}}{\sqrt{ab}} \left( y_1 - \frac{acd + bm}{ac^2 + b} \right), \]
then \( dy_1 = \frac{\sqrt{ab}}{\sqrt{2}\sqrt{ac^2 + b}} d\tilde{y}_1 \) and at \( y_1 = A, \tilde{y}_1 = \tilde{A} = \frac{\sqrt{2\sqrt{ac^2 + b}}}{\sqrt{ab}} (A - \frac{acd + bm}{ac^2 + b}). \)

Therefore,
\[ \int_{-\infty}^{A} \frac{1}{\sqrt{a\pi}} e^{-\frac{a^2}{a^2} (y_1 - \frac{a cd + bm}{ac^2 + b})^2} dy_1 = \frac{\sqrt{ab}}{\sqrt{2\sqrt{ac^2 + b}}} \int_{-\infty}^{\tilde{A}} e^{-\frac{\tilde{y}_1^2}{2}} d\tilde{y}_1 \]
\[ = \frac{\sqrt{b}}{\sqrt{2\sqrt{ac^2 + b}}} \Phi(\tilde{A}), \]
where $\tilde{A} = \frac{\sqrt{2 ac^2 + b}}{\sqrt{ab}} (A - \frac{ac_1 + bm}{ac^2 + b})$.

Returning to our original variables we obtain that

$$\tilde{A} = \frac{\sqrt{2a_1 (\frac{c_{11}}{c_{12}})^2 + 2a_2}}{\sqrt{2a_1 a_2}} \left[ \left( \frac{\ln K_1}{c_{11}} - \frac{c_{12} \tilde{m}_2 - (c_{12})^2}{c_{11}} \tilde{a}_2 \right) - \frac{2a_1 (\frac{c_{12}}{c_{11}})^2 - m_2 - c_{12} a_2}{2a_1 (\frac{c_{12}}{c_{11}})^2 + 2a_2} \right]$$

and

$$\frac{\sqrt{b}}{\sqrt{ac^2 + b}} = \frac{\sqrt{a_2}}{\sqrt{a_1 (\frac{c_{11}}{c_{12}})^2 + a_2}} \Phi(\tilde{A}) \sim O(a_2^{\frac{1}{4}}).$$

Therefore, as $a_2 \to 0$

$$\frac{\sqrt{a_2}}{\sqrt{a_1 (\frac{c_{11}}{c_{12}})^2 + a_2}} \Phi(\tilde{A}) \sim O(a_2^{\frac{1}{4}}).$$

Hence, there exits $C_5 > 0$ such that

$$S_1 \leq C_5 \sqrt{a_2}$$

for $a_2 > 0$ small enough. As a consequence, we obtain that there exists constants $C_1', C_2', C_3', C_4' > 0$ such that

$$|J_1 - J_1^0| \leq C_1' \sqrt{a_2} + C_2' (e^{C_3'a_2} - 1) + C_4' a_2$$

for $a_2 > 0$ small enough.

We have obtained an upper bound under the assumption of $c_{11}, c_{12} > 0$. In general, for the remaining cases we have that

- If $c_{11} > 0$ and $c_{12} < 0$ then

$$|J_1 - J_1^0| \leq \sum_{i=1}^{n} T_i,$$

with

$$T_1 = \frac{\sqrt{a_1}}{\sqrt{ac^2 + b}} \int_{-\infty}^{\frac{\ln K_1}{c_{12}} - \frac{c_{12} \tilde{m}_2 - (c_{12})^2}{c_{11}} \tilde{a}_2} \left( \frac{e^{c_{12} y_1} - \left( \frac{\ln K_1}{c_{12}} - (m_2 + c_{12} a_2) \right)}{\sqrt{a_2}} \right) p_{\alpha_1, \tilde{m}_1}(y_1) dy_1$$

and

$$T_2 = K_1 \int_{-\infty}^{\frac{\ln K_1}{c_{12}} - \frac{c_{12} \tilde{m}_2 - (c_{12})^2}{c_{11}} \tilde{a}_2} Q \left( \frac{e^{c_{12} y_1} - \left( \frac{\ln K_1}{c_{12}} - (m_2 + c_{12} a_2) \right)}{\sqrt{a_2}} \right) p_{\alpha_1, \tilde{m}_1}(y_1) dy_1$$
\[ T_3 = e^{c_{12}\tilde{m}_2} \left| 1 - e^{\frac{(c_{12})^2}{2}} \right| \int_{\ln \frac{K_1}{c_{11}}}^{\infty} s_1^1 e^{c_{11}y_1} p_{a_1, \tilde{m}_1}(y_1) dy_1 \]

\[ T_4 = s_1^1 e^{c_{12}\tilde{m}_2} e^{\frac{(c_{12})^2}{2}} \int_{\ln \frac{K_1}{c_{11}}}^{\infty} e^{c_{11}y_1} Q \left( \frac{\ln \frac{K_1}{c_{12}} - (\tilde{m}_2 + c_{12}\sigma_2)}{\sqrt{\sigma_2}} + \frac{c_{11}}{c_{12}} y_1 \right) p_{a_1, \tilde{m}_1}(y_1) dy_1 \]

\[ T_5 = K_1 \int_{\ln \frac{K_1}{c_{11}}}^{\infty} Q \left( \frac{\ln \frac{K_1}{c_{12}} - \tilde{m}_2}{\sqrt{\sigma_2}} + \frac{c_{11}}{c_{12}} y_1 \right) p_{a_1, \tilde{m}_1}(y_1) dy_1. \]

- If \( c_{11} < 0 \) and \( c_{12} > 0 \) then

\[ |J_1 - J_{10}^0| \leq \sum_{i=1}^{n} T_i, \]

with

\[ T_1 = s_1^1 e^{c_{12}\tilde{m}_2} e^{\frac{(c_{12})^2}{2}} \int_{\ln \frac{K_1}{c_{11}}}^{\ln \frac{K_1}{c_{12}}} e^{c_{11}y_1} Q \left( \frac{\ln \frac{K_1}{c_{12}} - (\tilde{m}_2 + c_{12}\sigma_2)}{\sqrt{\sigma_2}} + \frac{c_{11}}{c_{12}} y_1 \right) p_{a_1, \tilde{m}_1}(y_1) dy_1 \]

\[ T_2 = K_1 \int_{\ln \frac{K_1}{c_{11}}}^{\ln \frac{K_1}{c_{12}}} Q \left( \frac{\ln \frac{K_1}{c_{12}} - \tilde{m}_2}{\sqrt{\sigma_2}} + \frac{c_{11}}{c_{12}} y_1 \right) p_{a_1, \tilde{m}_1}(y_1) dy_1 \]

\[ T_3 = e^{c_{12}\tilde{m}_2} \left| 1 - e^{\frac{(c_{12})^2}{2}} \right| \int_{-\infty}^{\ln \frac{K_1}{c_{11}}} s_1^1 e^{c_{11}y_1} p_{a_1, \tilde{m}_1}(y_1) dy_1 \]

\[ T_4 = s_1^1 e^{c_{12}\tilde{m}_2} e^{\frac{(c_{12})^2}{2}} \int_{-\infty}^{\ln \frac{K_1}{c_{11}}} e^{c_{11}y_1} Q \left( \frac{c_{11}y_1 - (\ln \frac{K_1}{c_{12}} - \tilde{m}_2)}{\sqrt{\sigma_2}} \right) p_{a_1, \tilde{m}_1}(y_1) dy_1 \]

\[ T_5 = K_1 \int_{-\infty}^{\ln \frac{K_1}{c_{12}}} e^{c_{11}y_1} Q \left( \frac{c_{11}y_1 - (\ln \frac{K_1}{c_{12}} - \tilde{m}_2)}{\sqrt{\sigma_2}} \right) p_{a_1, \tilde{m}_1}(y_1) dy_1. \]

- If \( c_{11} < 0 \) and \( c_{12} < 0 \) then

\[ |J_1 - J_{10}^0| \leq \sum_{i=1}^{n} T_i, \]

with

\[ T_1 = s_1^1 e^{c_{12}\tilde{m}_2} e^{\frac{(c_{12})^2}{2}} \int_{\ln \frac{K_1}{c_{11}}}^{\ln \frac{K_1}{c_{12}}} e^{c_{11}y_1} Q \left( \frac{c_{11}y_1 - (\ln \frac{K_1}{c_{12}} - \tilde{m}_2)}{\sqrt{\sigma_2}} \right) p_{a_1, \tilde{m}_1}(y_1) dy_1 \]
We considered the problem of finding upper bounds of the magnitude of the change in price of a portfolio of options after eigenvalues from the covariance matrix are modified. Upper bounds were obtained using estimates of the \( L^1 \) distance of Gaussian densities assuming that the portfolio’s payoff is bounded. Upper bounds were also obtained for portfolios of \( n \) uncorrelated single asset options, two digital options with correlated underlying assets and two call options with correlated underlying assets.

\[
T_2 = K_1 \int_{-\infty}^{\infty} \frac{\ln K_1}{\sqrt{\pi}} e^{\frac{c_{12}^2 y_2}{2}} Q \left( \frac{\ln K_1}{\sqrt{\pi}} - \frac{c_{12}^2 y_2}{2} \right) p_{a_1, \tilde{m}_2} (y_1) dy_1
\]

\[
T_3 = e^{c_{12} \tilde{m}_2} \left| 1 - e^{\frac{(c_{12}^2/2)^i}{2}} \right| \int_{-\infty}^{\infty} e^{\frac{(c_{12}^2/2)^i}{2}} e^{c_{11} y_1} p_{a_1, \tilde{m}_1} (y_1) dy_1
\]

\[
T_4 = k_1 e^{c_{12} \tilde{m}_2} e^{\frac{(c_{12}^2/2)^i}{2}} \int_{-\infty}^{\infty} e^{\frac{(c_{12}^2/2)^i}{2}} e^{c_{11} y_1} Q \left( \frac{\ln K_1}{\sqrt{\pi}} - \frac{c_{12}^2 y_2}{2} \right) p_{a_1, \tilde{m}_1} (y_1) dy_1
\]

\[
T_5 = K_1 \int_{-\infty}^{\infty} e^{\frac{c_{11} \tilde{m}_2}{2}} e^{\frac{c_{12} \tilde{m}_2}{2}} e^{c_{11} y_1} Q \left( \frac{\ln K_1}{\sqrt{\pi}} - \frac{c_{12} \tilde{m}_2}{2} \right) p_{a_1, \tilde{m}_1} (y_1) dy_1.
\]

In the remaining cases, following the same procedure described above, we obtain that there exist constants \( C_1, C_2, C_3, C_4 > 0 \) such that

\[
|J_1 - J_1^0| \leq C_1 \sqrt{a_2} + C_2 (e^{a_2} - 1) + C_4 a_2
\]

for \( a_2 > 0 \) small enough.

Monte Carlo simulations were used to plot the magnitude of the difference between the price of a portfolio of 2 options with correlated underlying assets and the approximation obtained by making zero the smallest eigenvalue.

Figure 2.4 shows the plot for a portfolio of two digital options. 4000000 simulations were used. The following values were assigned to the parameters and variables of the options: \( S_i^0 = 100, K_i = 106 \) and \( D_i = 10 \), for \( i = 1, 2, \Delta = T - t = 1, \sigma_1 = 0.2, \sigma_2 = 0.3, \rho = 0.3 \) and \( r = 0.02 \).

Figure 2.5 shows the plot for a portfolio of two call options. 4000000 simulations were used. The following values were assigned to the parameters and variables of the options: \( S_i^0 = 100 \) and \( K_i = 94 \) for \( i = 1, 2, \Delta = T - t = 1, \sigma_1 = 0.2, \sigma_2 = 0.3, \rho = 0.3 \) and \( r = 0.02 \).

### 2.5 Conclusion

We considered the problem of finding upper bounds of the magnitude of the change in price of a portfolio of options after eigenvalues from the covariance matrix are modified. Upper bounds were obtained using estimates of the \( L^1 \) distance of Gaussian densities assuming that the portfolio’s payoff is bounded. Upper bounds were also obtained for portfolios of \( n \) uncorrelated single asset options, two digital options with correlated underlying assets and two call options with correlated underlying assets.
Figure 2.4: Magnitude of the difference between the price of a portfolio of 2 digital options and the approximation obtained by making the smallest eigenvalue zero.

Figure 2.5: Magnitude of the difference between the price of a portfolio of 2 call options and the approximation obtained by making the smallest eigenvalue zero.
Chapter 3

Gaussian versus non Gaussian factor models

3.1 Introduction

Multiple factor models decompose an asset’s return into factors common to all assets and a specific factor. They can be used to predict returns, to generate estimates of abnormal returns, and to describe the covariance structure of returns. The common factors are often interpreted as capturing fundamental risk components. The factor model isolates an asset’s sensitivity to these risk factors.

There are three main types of multifactor models for asset returns: macroeconomic factor models, fundamental factor models, and statistical factor models. Macroeconomic factor models use observable economic time series as measures of common factors in asset returns. Fundamental factor models use observable firm or asset specific attributes such as firm size and industry classification to determine common factors in asset returns. Statistical factor models treat the common factors as unobservable or latent factors. Zivot, E. and Wang, J. (2006) provide an excellent description of each type of model.

Each of the three types of multifactor models for asset returns has the general form:

\[
R_{i,t} = \alpha_i + \beta_{1,i}f_{1,t} + \beta_{2,i}f_{2,t} + \cdots + \beta_{K,i}f_{K,t} + \epsilon_{i,t}
\]

\[
= \alpha_i + \beta_i'f_t + \epsilon_{i,t},
\]

where \( R_{i,t} \) is the return (real or in excess of the risk-free rate) on asset \( i (i = 1, \cdots, N) \), \( f_{k,t} \) is the \( k \)-th common factor \( (k = 1, \cdots, K) \) and \( \epsilon_{i,t} \) is the asset specific factor in time period \( t (t = 1, \cdots, T) \), \( \alpha_i \) is the intercept and \( \beta_{k,i} \) is the factor loading for asset \( i \) on the \( k \)-th factor.

The multifactor model has the following assumptions:

- The factor realizations \( f_t \) are stationary with unconditional moments

\[
\mathbb{E}[f_t] = \mu_f \\
\text{Cov}(f_t) = \mathbb{E}[(f_t - \mu_f)'(f_t - \mu_f)] = \Omega_f.
\]
• The asset specific error terms $\epsilon_{i,t}$ are uncorrelated with each of the common factors $f_{k,t}$ so that

$$Cov(f_{k,t}, \epsilon_{i,t}) = 0, \text{ for all } k, i \text{ and } t.$$ 

• The error terms are serially uncorrelated and contemporaneously uncorrelated across assets

$$Cov(\epsilon_{i,t}, \epsilon_{j,s}) = \sigma^2_i \text{ for all } i = j \text{ and } t = s$$

$$= 0, \text{ otherwise}.$$ 

In a macroeconomic factor model, the realizations $f_t$ are observed macroeconomic variables that are assumed to be uncorrelated with the asset specific error terms $\epsilon_{i,t}$. The two most common macroeconomic factor models are Sharpe’s single factor model (1970) and Chen, Roll and Ross’s (1986) multifactor model. Because the macroeconomic factors are observed the problem is then to estimate the factor betas $\beta_{k,i}$, residual variances $\sigma^2_i$ and factor covariance $\Omega_f$, using time series regression techniques.

Sharpe’s single factor model or market model, is given by

$$R_{i,t} = \alpha_i + \beta_i R_{M,t} + \epsilon_{i,t}, \quad i = 1, \cdots, N \quad \text{and} \quad t = 1, \cdots, T,$$

where $R_{M,t}$ denotes the return or excess return (relative to the risk free rate) on a market index (typically a value weighted index like the S&P 500 index) in time period $t$. The market index is meant to capture market risk, while the error term captures non-market firm specific risk.

The general macroeconomic multifactor model specifies $K$ observable macro-economic variables as the factor realizations $f_t$:

$$R_i = 1_T \alpha_i + F \beta_i + \epsilon_i, \quad i = 1, \cdots, N$$

$$\mathbb{E}[^t \epsilon_i \epsilon_i^t] = \sigma^2_i I_T,$$

where $1_T$ is a $(T \times 1)$ vector of ones, $F$ is a $(T \times K)$ matrix of factor realizations (the $t$-th row of $F$ is $f_t^t$), $\beta_i$ is a $(K \times 1)$ vector of factor loadings, and $\epsilon_i$ is a $(T \times 1)$ vector of error terms with covariance matrix $\sigma^2_i I_T$.

The general form of the covariance matrix for the macroeconomic factor model is

$$\Omega = B \Omega_f B^T + D,$$

where $B = [\beta_1, \beta_2, \cdots, \beta_N]^T$, $\Omega_f = \mathbb{E}[(f_t - \mu_t)(f_t - \mu_t)^t]$ is the covariance matrix of the observed factors and $D$ is a diagonal matrix with $\sigma^2_i = \text{Var}(\epsilon_{i,t})$ along the diagonal.

Because the factor realizations are observable, the parameter matrices $B$ and $D$ of the model may be estimated using time series regression. The covariance matrix of the factor realizations may be estimated using the time series sample covariance matrix.

Fundamental factor models use observable asset specific characteristics (fundamentals) like industry classification, market capitalization, style classification (value, growth) to determine the common risk factors. The two most popular ways to estimate the fundamental factors are the ‘BARRA’ approach introduced in 1975 by Bar Rosenberg, founder of BARRA Inc., which is discussed in detail in Grinold and Kahn (2000) and the ‘Fama-French’ approach proposed by Fama, E and French, K. (1992).

In the BARRA, approach the observable asset specific fundamentals are treated as the factor betas $\beta_i$ which are time invariant, while the factor realizations at time $t$, $f_t$, are unobservable. The problem is then to estimate the factor realizations at time $t$ given the factor betas, which is done running $T$ cross-section regressions.
The BARRA-type single factor model is given by

\[ R_t = \beta f_t + \epsilon_t, \quad t = 1, \cdots, T, \]

where \( \beta \) is a vector of observed values of an asset specific attribute (market capitalization, industry classification, style classification) and \( f_t \) is an unobserved factor realization. It is assumed that

\[
\begin{align*}
\text{Var}(f_t) &= \sigma_f^2 \\
\text{Cov}(f_t, \epsilon_{i,t}) &= 0, \text{ for all } i, t \\
\text{Var}(\epsilon_{i,t}) &= \sigma_i^2, i = 1, \cdots, N.
\end{align*}
\]

The factor realization \( f_t \) is the parameter to be estimated for each time period \( t = 1, \cdots, T \). Because the error term \( \epsilon_t \) is heteroskedastic, efficient estimation of \( f_t \) is done by weighted least squares given that the specific variances \( \sigma_i^2 \) are known, in fact

\[
\hat{f}_{t, wls} = (\beta' D^{-1} \beta)^{-1} \beta' D^{-1} R_t, \quad t = 1, \cdots, T,
\]

where \( D \) is a diagonal matrix with \( \sigma_i^2 \) along the diagonal.

A popular fundamental factor model with \( K \) factors is the following stylized BARRA-type industry factor model with \( K \) mutually exclusive industries

\[
\begin{align*}
R_t &= \beta_1 f_{1,t} + \cdots + \beta_K f_{K,t} + \epsilon_t, \quad t = 1, \cdots, T \\
&= B f_t + \epsilon_t, \\
\mathbb{E}[\epsilon_t \epsilon_t'] &= D \\
\text{Cov}(f_t) &= \Omega_f,
\end{align*}
\]

where \( R_t \) is an \((N \times 1)\) vector of returns, \( B = [\beta_1, \cdots, \beta_K] \) is a \((N \times K)\) matrix of zeros and ones that reflects the industry factor sensitivities for each asset, \( f_t = (f_{1,t}, \cdots, f_{K,t})' \) is a \((K \times 1)\) vector of unobserved factor realizations, \( \epsilon_t \) is an \((N \times 1)\) error term, and \( D \) is the diagonal matrix with \( \sigma_i^2 \) along the diagonal.

The factor realizations are estimated as follows:

- For each \( t = 1, \cdots, T \) an ordinary linear regression estimate is obtained,

\[
\hat{f}_{t, OLS} = (B' B)^{-1} B' R_t.
\]

- Given the time series of factor realizations \( \hat{f}_{OLS} = (\hat{f}_{1, OLS}, \cdots, \hat{f}_{T, OLS}) \), the covariance matrix of the industry factors is estimated as the time series sample covariance

\[
\hat{\Omega}_f^{OLS} = \frac{1}{T-1} \sum_{t=1}^{T} (\hat{f}_{t, OLS} - \bar{f}_{OLS})(\hat{f}_{t, OLS} - \bar{f}_{OLS})' \\
\bar{f}_{OLS} = \frac{1}{T} \sum_{t=1}^{T} \hat{f}_{t, OLS}.
\]

Similarly, the residual variances can be estimated using

\[
\hat{\sigma}_i^{2, OLS} = \frac{1}{T-1} \sum_{t=1}^{T} (\hat{\epsilon}_{i,t, OLS} - \bar{\epsilon}_{i, OLS})^2, \quad i = 1, \cdots, N. \\
\bar{\epsilon}_{i, OLS} = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_{i,t, OLS}.
\]
The covariance matrix of the $N$ assets is then estimated by

$$\hat{\Omega}_{OLS} = B\hat{\Omega}_{OLS}^{F}B' + \hat{D}_{OLS},$$

where $\hat{D}_{OLS}$ is a diagonal matrix with $\hat{\sigma}_{i,OLS}^2$ along the diagonal.

- The estimates of the residual variances are used as weights for weighted least squares estimation of the factor realizations

$$\hat{f}_{t, GLS} = (B'\hat{\Omega}_{OLS}^{-1}\hat{D}_{OLS}^{-1}B')^{-1}B'\hat{\Omega}_{OLS}^{-1}B'(R_t - \bar{R}_t1_N).$$

Using the time series of factor realizations $(\hat{f}_{1, GLS}, \ldots, \hat{f}_{T, GLS})$, the covariance matrix of the industry factors $\hat{\Omega}_{GLS}^{F}$ and the residual variances $\hat{\sigma}_{i, GLS}^2$ are estimated. Then the covariance matrix of the $N$ assets is re-estimated by

$$\hat{\Omega}_{GLS} = B\hat{\Omega}_{GLS}^{F}B' + \hat{D}_{GLS},$$

where $\hat{D}_{GLS}$ is a diagonal matrix with $\hat{\sigma}_{i, GLS}^2$ along the diagonal.

In the Fama-French approach, the factor realizations are determined using a two step procedure, given observed asset characteristics. First, the cross section of assets is sorted based on the values of the asset specific characteristics. Then a hedge portfolio is constructed that is long in the top quintile and short in the bottom quintile of the sorted assets. The observed return of this hedge portfolio at time $t$ is the observed factor realization for the asset specific characteristic. The process is repeated for each asset specific characteristic. Then, given the observed factor realizations for $t = 1, \ldots, T$ the factor betas for each asset are estimated using $N$ time series regressions.

In statistical factor models the factor realizations $f_t$ are not directly observable and most be extracted from the observable returns $R_t$ using statistical methods. The two main methods are factor analysis and principal components.

The traditional factor analysis assumes that

$$R_t = \mu + Bf_t + \epsilon_t$$

$$Cov(f_t, \epsilon_s) = 0, \text{ for all } t, s$$

$$\mathbb{E}[f_t] = 0$$

$$\mathbb{E}[\epsilon_t] = 0$$

$$Var(f_t) = I_K$$

$$Var(\epsilon_t) = D,$$

where $D$ is a diagonal matrix with $\sigma_t^2$ along the diagonal. Assuming that returns are jointly normally distributed and temporarily i.i.d. $B$ and $D$ can be estimated using maximum likelihood. Given estimates $\hat{B}$ and $\hat{D}$, an empirical version of the factor model can be constructed as

$$R_t - \hat{\mu} = \hat{B}f_t + \hat{\epsilon}_t,$$

where $\hat{\mu}$ is the sample mean vector of $R_t$. The factor realizations in a given time period $t$, $f_t$, are then estimated using the cross sectional generalized least squares regression,

$$\hat{f}_t = (\hat{B}'\hat{D}^{-1}\hat{B})^{-1}\hat{B}'\hat{D}^{-1}(R_t - \hat{\mu})$$
and the estimated factor model covariance matrix is given by

$$\hat{\Omega}^F = \hat{B} \hat{B}' + \hat{D}. $$

Principal components analysis is used to explain the majority of the information in the sample covariance matrix of returns. With $N$ assets there are $N$ principal components, and these principal components are just linear combinations of returns. The principal components are constructed and ordered so that the first principal component explains the largest portion of the sample covariance matrix of returns, the second principal component explains the next largest portion, and so on. The principal components are constructed to be orthogonal to each other and to be normalized to have unit length. The $K$ most important principal components are the factor realizations. The factor loadings are then estimated using regression techniques.

Macroeconomic factor models and in particular Sharpe’s single index model assume time invariant parameters. This assumption may be inadequate, because a constant estimate often fails to capture the dynamic relation between an individual investment and the market. A simple and applicable technique to account for dynamic changes in the model’s parameters is a rolling window ordinary least squares regression. The intercept $\alpha_t$ and the factor loading $\beta_t$ at time $t$ are determined by regressing return series from $t - \tau$ to $t - 1$. The determination of the window’s length, $\tau$, is somewhat ad hoc.

We consider four 1-factor models with time varying parameters. Each of the considered models has the same general structure. For each $t \geq \tau + 1$,

$$R_{f,t-j} = \alpha_t + \beta_t R_{m,t-j} + \epsilon_{t-j}, \quad \text{for all } j = 1, \ldots, \tau, \quad (3.1)$$

$R_{f,t-j} = r_{f,t-j} - r_{F,t-j}$ and $R_{m,t-j} = r_{m,t-j} - r_{F,t-j}$ where $r_{f,t-j}$, $r_{m,t-j}$ and $r_{F,t-j}$ denote the daily return of an asset, the daily return of the market and the daily risk free rate respectively, at time $t - j$.

The differences among the considered models lie on the assumptions on the model components. Model 1 assumes that the residuals $(\epsilon_{t-j})$ are i.i.d. such that $\epsilon_{t-j} \sim N(0, \sigma_{\epsilon,t}^2)$ for all $j = t - \tau, \ldots, t - 1$, which is a standard assumption of factor models. The other models combine the structure of a 1-factor model with time varying coefficients, with dynamic volatility assumptions on the model components. Model 2 assumes that the series of residuals $(\epsilon_{t-j})$ follows a GARCH(1,1) process. Model 3 assumes that the series of centered excess market returns $(R_{m,t-j} - \mu_{m,t})$ follows a GARCH(1,1) process and the series of residuals $(\epsilon_{t-j})$ is i.i.d. such that $\epsilon_{t-j} \sim N(0, \sigma_{\epsilon,t}^2)$ for all $j = t - \tau, \ldots, t - 1$. Model 4 assumes that both the series of centered excess asset returns and the series of centered excess market returns follow GARCH(1,1) processes which leads us to consider weak GARCH(2,2) residuals $(\epsilon_{t-j})$ for this model. The four models previously mentioned are then used to estimate the time varying alpha and beta of three different hedge fund strategies.

This chapter starts with a description of the assumptions, and most common estimation methods of the linear regression model. In section 3.3 basic concepts and results from the theory of time series are introduced. Section 3.4 focuses on the GARCH model and its estimation methods. In section 3.5 the four 1-factor models mentioned above and their main assumptions and estimation methods are discussed. Section 3.6 shows an application of these models to hedge fund return modelling, in particular, the performance and market exposure of three different hedge fund strategies are estimated and results are compared. Section 3.7 concludes this chapter.
3.2 Linear regression and estimation methods

Linear regression plays a key role in factor modeling. In fact, rolling window linear regression has become the standard method to estimate dynamic alpha and betas in macroeconomic factor models. This section provides a detailed description of the simple linear regression model, its main assumptions, properties, and estimation methods.

**Definition 3.2.1.** The simple linear regression model is given by the equation:

\[ y_t = \beta_1 + \beta_2 x_t + \epsilon_t \]  

where \((\epsilon_t)_{t=1}^n\) are independent and identically distributed Gaussian random variables with mean 0 and variance \(\sigma^2\).

In general, we want to estimate the parameters \(\beta_1\) and \(\beta_2\) given a set of observations of a dependent variable \(y_t\) and a single independent variable \(x_t\). The total number of observations is called the sample size and is denoted by \(n\). In practice, when the successive observations are ordered in time, it often seem plausible that an error term will be correlated with neighbouring error terms. This phenomenon is called serial correlation, when there is serial correlation, the error terms cannot be i.i.d. because they are not independent. Another possibility is that the variance of the error terms may be systematically larger for some observations than for others, which may indicate the presence of conditional heteroskedasticity. Later on we discuss models that capture these types of behavior, the ARMA and GARCH models.

The simple linear regression can be written in matrix form:

\[ y = X\beta + \epsilon \]

with

\[
\begin{align*}
  y &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \\
  \epsilon &= \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \\
  X &= \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots \\ 1 & x_n \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}
\end{align*}
\]

Different methods can be used to estimate the model parameters \(\beta_1\) and \(\beta_2\). The most popular ones are least squares and maximum likelihood estimation.

3.2.1 Least squares estimation

Consider the expression \(y_t - X_t\beta\), if \(\beta\) is allowed to vary arbitrarily, then this difference is called the residual associated with the \(t\)-th observation. Similarly the \(n\)-dimensional vector \(y - X\beta\) is called the vector of residuals.

The sum of squares of the components of the vector of residuals is called the sum of squared residuals, or SSR. The idea of least squares estimation is to minimize the sum of squared residuals associated with a regression model. In particular for the ordinary linear regression we have that:

\[ SSR(\beta) = \sum_{t=1}^{n} (y_t - X_t\beta)^2 \]
which can be rewritten in matrix form as

\[ SSR(\beta) = (y - X\beta)'(y - X\beta). \]

**Proposition 28.** The least squares estimator of \( \beta \) from the simple linear regression model is

\[ \hat{\beta} = (X'X)^{-1}X'y. \]

**Proof.** We must find a vector \( \beta \) that minimizes \( SSR(\beta) = (y - X\beta)'(y - X\beta) \),

\[
SSR(\beta) = (y - X\beta)'(y - X\beta) \\
= (y' - (X\beta)')(y - X\beta) \\
= y'y - y'(X\beta) - (X\beta)'y + (X\beta)'(X\beta).
\]

Therefore,

\[
\frac{\partial SSR}{\partial \beta} = -2y'X + 2\beta'(X'X).
\]

Hence,

\[
\frac{\partial SSR}{\partial \beta} = 0 \iff \beta'(X'X) - y'X = 0 \\
\iff (X'X)\beta - X'y = 0.
\]

If \( X \) is full rank then \( X'X \) is positive definite and

\[ \beta = (X'X)^{-1}X'y. \]

\[ \square \]

Let’s consider some statistical properties that are desirable for an estimator to possess.

**Definition 3.2.2.** Suppose that \( \hat{\theta} \) is an estimator of some parameter \( \theta \), the true value of which is \( \theta_0 \), then the bias of \( \hat{\theta} \) is defined as \( \mathbb{E}(\hat{\theta}) - \theta_0 \). If the bias of an estimator is zero for every admissible value of \( \theta_0 \), then the estimator is said to be unbiased.

If the matrix \( X \) is considered fixed or non-stochastic then the least squares estimator of the OLR \( \hat{\beta} \) is unbiased. However this assumption is often too strong, so usually it is assumed that \( \epsilon \) and \( X \) are independent, or the even weaker exogeneity assumption is made

\[
\mathbb{E}(\epsilon|X) = 0 \quad (3.3)
\]

to show that \( \hat{\beta} \) is unbiased.

**Theorem 29.** If it is assumed that \( \mathbb{E}(\epsilon|X) = 0 \) then the ordinary least squares estimator \( \hat{\beta} \) is unbiased.

**Proof.** Let \( \beta_0 \) be a vector with the true parameters of model (3.2). We have that:

\[
\hat{\beta} = (X'X)^{-1}X'y \\
= (X'X)^{-1}X'(X\beta_0 + \epsilon) \\
= \beta_0 + (X'X)^{-1}X'\epsilon.
\]
Therefore,
\[
E(\hat{\beta}) = \beta_0 + E((X'X)^{-1}X'\epsilon) \\
= \beta_0 + E(E((X'X)^{-1}X'\epsilon|X)) \\
= \beta_0 + E((X'X)^{-1}X'E(\epsilon|X)) \\
= \beta_0.
\]

\[\square\]

**Definition 3.2.3.** Let \(a(y^n)\) be a vector function of a random vector \(y^n\), \(a(y^n)\) converges in probability to a random vector \(a_0\) if, for all \(\epsilon > 0\),
\[
\lim_{n \to \infty} P(||a(y^n) - a_0|| < \epsilon) = 1.
\]
We denote the probability limit by \(\text{plim}\), so \(\text{plim}_{n \to \infty} a(y^n) = a_0\) means that \(a(y^n)\) converges in probability to \(a_0\).

**Theorem 30** (Law of large numbers). Let \(\bar{x}\) be the sample mean of \(x_t\), with \(t = 1, \cdots, n\), a sequence of random variables, each with expectation \(\mu\). Then, provided the \(x_t\) are independent, a law of large numbers states that
\[
\text{plim}_{n \to \infty} \bar{x} = \text{plim}_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} x_t = \mu
\]

**Proof.** See Rosenthal, J. (2006), p. 60. \[\square\]

**Definition 3.2.4.** A **consistent** estimator is one for which the estimate converges (in probability) to the quantity being estimated as the size of the sample tends to infinity.

**Theorem 31.** If \(\text{plim}_{n \to \infty} X'X = S_{X'X}\) where \(S_{X'X}\) is a non-stochastic matrix with full rank and the (exogeneity) condition (3.3) is assumed then the ordinary least squares estimator \(\hat{\beta}\) is consistent.

**Proof.** When proving Theorem (29), we showed that
\[
\hat{\beta} = \beta_0 + (X'X)^{-1}X'\epsilon.
\]
Therefore,
\[
\text{plim}_{n \to \infty} \hat{\beta} = \beta_0 + \text{plim}_{n \to \infty} (X'X)^{-1}X'\epsilon \\
= \beta_0 + (\text{plim}_{n \to \infty} \frac{1}{n} X'X)^{-1} \text{plim}_{n \to \infty} \frac{1}{n} X'\epsilon \\
= \beta_0 + (S_{X'X})^{-1} \text{plim}_{n \to \infty} \frac{1}{n} X'\epsilon.
\]
However, from Theorem (30) (Law of large numbers) and the exogeneity condition, we obtain that
\[
\text{plim}_{n \to \infty} \frac{1}{n} X'\epsilon = \text{plim}_{n \to \infty} \sum_{t=1}^{n} X'_t \epsilon_t \\
= E(X'_t \epsilon_t) \\
= 0.
\]
Note that
\[
E(X'_t \epsilon_t) = E(E(X'_t \epsilon_t|X_t)) \\
= E(X'_t E(\epsilon_t|X_t)) \\
= 0
\]
because \(E(\epsilon_t|X_t) = 0\). \[\square\]
**Definition 3.2.5.** One estimator is said to be more efficient than another if, on average the former yields more accurate estimates than the latter.

**Theorem 3.2 (Gauss-Markov Theorem).** If it is assumed that \( E(\epsilon|X) = 0 \) and \( E(\epsilon\epsilon'|X) = \sigma^2I \) in the linear regression model (3.2), then the ordinary least squares estimator \( \hat{\beta} \) is more efficient than any other linear unbiased estimator \( \tilde{\beta} \).


According to Theorem 3.2 (Gauss-Markov), if the parameters of a regression model are to be estimated efficiently by least squares, the error terms must be uncorrelated and have the same variance: \( E(\epsilon\epsilon') = \sigma^2I \).

However, regression models can have error terms that are heteroskedastic, serially correlated, or both. Consider the generalized linear regression model:

\[
y = X\beta + \epsilon, \quad E(\epsilon\epsilon') = \Omega,
\]

where \( \Omega \), the covariance matrix of the error terms, is a positive definite \( n \times n \) matrix.

If \( \Omega \) is equal to \( \sigma^2I \), it is just the linear regression model with uncorrelated and homoskedastic errors. If \( \Omega \) is diagonal with non constant diagonal elements, then the errors are still uncorrelated, but they are heteroskedastic. If \( \Omega \) is not diagonal, then \( \epsilon_i \) and \( \epsilon_j \) are correlated whenever \( \Omega_{ij} \) the \( ij^{th} \) element of \( \Omega \), is non zero.

An efficient estimator for this model can be obtained transforming the regression so that it satisfies the conditions of Theorem 3.2 (Gauss-Markov). This transformation is expressed in terms of an \( n \times n \) matrix \( \Psi \), which is usually triangular, that satisfies the equation

\[
\Omega^{-1} = \Psi\Psi'.
\]

Multiplying the generalized linear regression equation by \( \Psi \) gives

\[
\Psi' y = \Psi'X\beta + \Psi'\epsilon.
\]

In addition,

\[
E(\Psi'\epsilon\epsilon'\Psi) = \Psi'E(\epsilon\epsilon')\Psi = \Psi'\Omega\Psi = \Psi'(\Psi\Psi')^{-1}\Psi = \Psi'(\Psi')^{-1}(\Psi)^{-1}\Psi = I.
\]

Therefore, we obtain that

\[
\Psi' y = \Psi'X\beta + \Psi'\epsilon, \quad E(\Psi'\epsilon\epsilon'\Psi) = I.
\]

Because the covariance matrix \( \Omega \) is non singular, the matrix \( \Psi \) must be as well, and so the transformed regression model (3.6) is equivalent to the original model (3.4).

The ordinary least squares estimator of \( \beta \) from regression model (3.6) is

\[
\hat{\beta}_{GLS} = (X'\Psi\Psi'X)^{-1}X'\Psi\Psi'y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.
\]
This estimator is called the *generalized least squares estimator* of $\beta$. Because $\Psi$ is constant or non deterministic under the assumptions of Theorems (29), (31) and (32), the estimator $\hat{\beta}_{\text{GLS}}$ is unbiased, consistent and efficient respectively.

The generalized least squares estimator $\hat{\beta}_{\text{GLS}}$ can also be obtained by minimizing the generalized least squares criterion function

$$(y - X\beta)'\Omega^{-1}(y - X\beta),$$

which is the sum of squared residuals from the transformed regression (3.6). This criterion function can be thought of as a generalization of the $\text{SSR}$ function in which the squares and cross products of the residuals from the original equation (3.4) are weighted by the inverse of the matrix $\Omega$. When $\Omega$ is a diagonal matrix, each observation is simply given a weight proportional to the inverse of the variance of its error term. This is the case when the errors terms are heteroskedastic but uncorrelated.

Let $\omega^2_t$ denote the $t^{th}$ diagonal element of $\Omega$. Then $\Omega^{-1}$ is a diagonal matrix with $t^{th}$ diagonal element $\omega^{-2}_t$, and $\Psi$ can be chosen as the diagonal matrix with $t^{th}$ diagonal element $\omega^{-1}_t$. Therefore, for $t = 1, \cdots, n$ we can write

$$\omega^{-1}_t y_t = \omega^{-1}_t X_t \beta + \omega^{-1}_t \epsilon_t.$$

This regression can be estimated by ordinary least squares. This special case of generalized least squares is often called *weighted least squares*. Note that observations for which the variance of the error terms are large, are given low weights, and observations for which it is small are given high weights.

**Remark.** If the vector of regression functions was $x(\beta)$ instead of $X\beta$ we can obtain generalized non-linear least squares estimates by minimizing the criterion function

$$(y - x(\beta))'\Omega^{-1}(y - x(\beta)).$$

If we differentiate equation (3.8) with respect to $\beta$ and divide the result by $-2$, we obtain that

$$X'(\beta)\Omega^{-1}(y - x(\beta)) = 0,$$

where $X(\beta)$ is the matrix of derivatives of $x(\beta)$ with respect to $\beta$. Solving equation (3.9) requires some sort of minimization procedure.

In practice, the covariance matrix $\Omega$ is often not known. However, in many cases it is reasonable to suppose that $\Omega$ depends in a known way on a vector of unknown parameters $\gamma$. If so, it may be possible to estimate $\gamma$ consistently, so as to obtain $\Omega(\hat{\gamma})$. Then $\Psi(\hat{\gamma})$ can be defined as in equation (3.5) and generalized least squares estimates computed conditional on $\Psi(\hat{\gamma})$. This type of procedure is called *feasible generalized least squares*. Under suitable regularity conditions it can be shown that this type of procedure yields a feasible generalized least squares estimator $\hat{\beta}_F$ that is consistent and asymptotically equivalent to the generalized least squares estimator $\hat{\beta}_{\text{GLS}}$, see Amemiya, T. (1973).

### 3.2.2 Maximum likelihood estimation

Let $y$ be an $n$-dimensional vector of observations of a random variable, for a given $k$-vector $\theta$ of parameters, let the joint density function of $y$ be written as $f(y, \theta)$. If $f(y, \theta)$ is evaluated at the $n$-dimensional $y$ found in a given data set, then the function $f(y, \cdot)$ of the model parameters is referred to as the *likelihood function* of the model for a given data set. The *maximum likelihood* estimation method maximizes the loglikelihood function with respect
to the parameters. A parameter vector $\hat{\theta}$ at which the likelihood takes on its maximum value is called a maximum likelihood estimate of the parameters.

Often, the successive observations in a sample are assumed to be independent. In that case the joint density of the entire sample is the product of the densities of the observations. Let $f(y_t, \theta)$ denote the probability density function of the $t^{th}$ observation. Then the joint density of the entire sample $y$ is

$$f(y, \theta) = \prod_{t=1}^{n} f(y_t, \theta). \quad (3.10)$$

However, it is customary to work instead with the log-likelihood function $L(y, \theta) \equiv \log f(y, \theta) = \sum_{t=1}^{n} L_t(y, \theta)$,

where $L_t(y, \theta)$, the contribution to the log-likelihood function made by observation $t$, is equal to $\log f_t(y_t, \theta)$. Whatever value of $\theta$ that maximizes the log-likelihood function (3.11) will also maximize the likelihood function (3.10) because $L(y, \theta)$ is just a monotonic transformation of $f(y, \theta)$.

For regression models, if we make the assumption that the error terms are normally distributed, the maximum likelihood estimators coincide with some of the least squares estimators that we mentioned previously. Maximum likelihood estimation can be applied to a wide variety of models other than regression models, yielding estimators with excellent asymptotic properties. However, its main disadvantage is that it requires stronger distributional assumptions than other estimation methods.

Consider the classical normal linear model

$$y = X\beta + u, \quad u \sim N(0, \sigma^2 I). \quad (3.12)$$

For this model, the explanatory variables in the matrix $X$ are assumed to be exogenous. Therefore, to construct the likelihood function, we can use the density of $y$ conditional on $X$. Note that $y_t$ is distributed, conditionally on $X$ as $N(X_t\beta, \sigma^2)$, so the probability density function of $y_t$ is

$$f_t(y_t, \beta, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(y_t - X_t\beta)^2}{2\sigma^2}\right).$$

The contribution to the log-likelihood function made by the $t^{th}$ observation is

$$L_t(y_t, \beta, \sigma) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y_t - X_t\beta)^2.$$

Because the $y_t$ are independent, the log-likelihood is the sum of these contributions over all $t$:

$$L(y, \beta, \sigma) = \frac{-n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^{n} (y_t - X_t\beta)^2.$$

To find the Maximum Likelihood estimator, we need to maximize (3.13)

**Theorem 33.** The Maximum Likelihood estimator of $\beta$ coincides with the ordinary least squares estimator under the assumptions of the classical normal linear model.

**Proof.** To maximize $L(y, \beta, \sigma)$, we first differentiate (3.13) with respect to $\sigma$, then we solve the resulting first
order condition for $\sigma$ as a function of the data and the remaining parameters, and then substitute back into (3.13).

Differentiating (3.13) with respect to $\sigma$ and equating the derivative to zero, yields the first order equation

$$\frac{\partial L(y, \beta, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} (y - X\beta)'(y - X\beta) = 0,$$

(3.14)

and solving (3.14) yields the result that

$$\hat{\sigma}^2(\beta) = \frac{1}{n} (y - X\beta)'(y - X\beta).$$

Substituting $\hat{\sigma}^2(\beta)$ into equation (3.13), we obtain the concentrated likelihood function

$$L^c(y, \beta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \left(\frac{1}{n} (y - X\beta)'(y - X\beta)\right) - \frac{n}{2}.$$

Maximizing the concentrated log-likelihood function is equivalent to minimizing the sum of squared residuals as function of $\beta$. Therefore the maximum likelihood estimator $\hat{\beta}$ must be identical to the ordinary least squares estimator.

Remark. Note that the fact that the maximum likelihood estimator and the ordinary least squares estimator are identical depends critically on the assumption that error terms are normally distributed. A different distribution would have yield a different maximum likelihood estimator.

One of the attractive features of the maximum likelihood estimation is that the estimators are consistent under quite weak regularity conditions and asymptotically normally distributed under stronger conditions.

**Theorem 3.4.** Let $L(y, \theta)$ denote a likelihood function, then under the following assumptions:

1. If $L(y, \theta_1) = L(y, \theta_2)$ then $\theta_1 = \theta_2$ for all $y$ (finite sample identification condition)

2. If $\theta_0$ is the true model parameter then $\lim n^{-1}L(y, \theta^*) \neq \lim n^{-1}L(y, \theta_0)$ for all $\theta^* \neq \theta_0$ (asymptotic identification condition)

The Maximum Likelihood estimator $\hat{\theta}_{MLE}$ is consistent.

Proof. Let $L(\theta) = \exp(L(\theta))$ denote the likelihood function, note that the dependency on $y$ of both $L$ and $L$ has been suppressed for notational simplicity. Suppose that $\theta_0$ is the true parameter vector and $\theta^*$ another vector in the parameter space of the model.

Since the logarithm function is strictly concave over the nonnegative real line and log-likelihood functions are nonnegative, using *Jensen’s Inequality*: If $X$ is a real valued random vector, then $E(h(X)) \leq h(E(X))$ whenever $h(\cdot)$ is concave; we obtain that

$$E_0 \log \left(\frac{L(\theta^*)}{L(\theta_0)}\right) < \log E_0 \left(\frac{L(\theta^*)}{L(\theta_0)}\right).$$

Note that this inequality holds strictly for all $\theta^* \neq \theta_0$ because of our first assumption. In addition, $E_0(\cdot)$ denotes the expectation taken with respect to the true parameter $\theta_0$. Since the joint density of the sample is the likelihood function evaluated at $\theta_0$, we have that

$$E_0 \left(\frac{L(\theta^*)}{L(\theta_0)}\right) = \int \frac{L(\theta^*)}{L(\theta_0)} L(\theta_0)dy = \int L(\theta^*)dy = 1.$$
Therefore,
\[
\log E_0 \left( \frac{L(\theta^*)}{L(\theta_0)} \right) = 0.
\]
As a consequence,
\[
E_0 \log \left( \frac{L(\theta^*)}{L(\theta_0)} \right) = E_0 \mathcal{L}(\theta^*) - E_0 \mathcal{L}(\theta_0) < 0. \tag{3.15}
\]
Applying a Law of Large Numbers to the contributions to the log-likelihood function, then
\[
p \lim n^{-1} \mathcal{L}(\theta) = \lim n^{-1} E_0 \mathcal{L}(\theta).
\]
As a consequence, from (3.15)
\[
p \lim_{n \to \infty} \frac{1}{n} \mathcal{L}(\theta^*) \leq p \lim_{n \to \infty} \frac{1}{n} \mathcal{L}(\theta_0).
\]
However, since \( \hat{\theta}_{MLE} \) maximizes \( \mathcal{L}(\theta) \)
\[
p \lim_{n \to \infty} \frac{1}{n} \mathcal{L}(\hat{\theta}_{MLE}) \geq p \lim_{n \to \infty} \frac{1}{n} \mathcal{L}(\theta_0).
\]
Therefore,
\[
p \lim_{n \to \infty} \frac{1}{n} \mathcal{L}(\hat{\theta}_{MLE}) = p \lim_{n \to \infty} \frac{1}{n} \mathcal{L}(\theta_0).
\]
This result on its own does not prove that \( \hat{\theta}_{MLE} \) is consistent, because there may be other vectors \( \theta^* \) for which
\[
p \lim_{n \to \infty} \frac{1}{n} \mathcal{L}(\theta^*) = p \lim_{n \to \infty} \frac{1}{n} \mathcal{L}(\theta_0),
\]
however our second assumption rules out this possibility, hence the maximum likelihood estimator \( \hat{\theta}_{MLE} \) is consistent.

Maximum Likelihood estimators have other attractive features, under certain regularity assumptions they are asymptotically normally distributed and asymptotically more efficient than Least Squares estimators, however they are not unbiased in general. See Davidson, R. and MacKinnon, J. (2004).

Let’s consider next the structure of the likelihood and log-likelihood functions when successive observations are not independent. Suppose that \( y_1, y_2 \) and \( y_3 \) are random variables defined in the same probability space then
\[
f(y_1, y_2, y_3) = f(y_1, y_2)f(y_3|y_1, y_2).
\]
However,
\[
f(y_1, y_2) = f(y_1)f(y_2|y_1).
\]
Therefore,
\[
f(y_1, y_2, y_3) = f(y_1)f(y_2|y_1)f(y_3|y_1, y_2).
\]
In general, for a sample of size \( n \),
\[
f(y_1, y_2, \ldots, y_n) = f(y_1)f(y_2|y_1) \cdots f(y_n|y_1, \ldots, y_{n-1}).
\]
This result can be rewritten as
\[
f(y^n) = \prod_{t=1}^{n} f(y_t|y^{t-1}),
\]
where \( y^t \) is a vector with components \( y_1, y_2, \ldots, y_t \). The vector \( y^t \) can be thought as a subsample consisting of the first \( t \) observations of the full sample.
For a model to be estimated by maximum likelihood, the density \( f(y^n) \) will depend on \( k \)-dimensional vector of parameters \( \theta \), and we can then write
\[
f(y^n, \theta) = \prod_{t=1}^{n} f(y_t | y_{t-1}; \theta). \tag{3.16}
\]
The structure of (3.16) is a straightforward generalization of that of (3.10) with the marginal densities of successive observations replaced by densities conditional on the previous observations.

The log-likelihood function corresponding to (3.16) has an additive structure
\[
\mathcal{L}(y, \theta) \equiv \log f(y^n, \theta) = \sum_{t=1}^{n} \mathcal{L}_t(y_t, \theta), \tag{3.17}
\]
where \( \mathcal{L}_t(y_t, \theta) \) is the contribution to the log-likelihood function made by observation \( t \). Note that, in the contributions \( \mathcal{L}_t(\cdot) \) to the log-likelihood we do not distinguish between current variable \( y_t \) and the lag variables in the vector \( y_{t-1} \). Therefore, (3.17) has exactly the same structure as (3.11).

### 3.3 Time series and stationary processes

The standard time series analysis rests on important concepts such as stationarity, white noise, innovation, autocorrelation, and on a central family of models, the autoregressive moving average (ARMA) models.

**Definition 3.3.1.** A sequence of random variables \((X_t)_{t \in \mathbb{Z}}\) defined on the same probability space is called a **time series**, and is an example of a discrete stochastic process.

Stationarity plays a central part in time series analysis, because it replaces in a natural way the assumption of independent and identically distributed (i.i.d.) distributions in standard statistics, which is often unrealistic in practice.

Next, we introduce two standard notions of stationarity.

**Definition 3.3.2.** The process \((X_t)\) is said to be **strictly (or strongly) stationary** if the vectors \((X_1, \cdots, X_k)\) and \((X_{1+h}, \cdots, X_{k+h})\) have the same joint distribution, for any \( k \in \mathbb{N} \) and \( h \in \mathbb{Z} \).

The following notion may seem less demanding because it only constraints the first two moments of the variables \( X_t \), but contrary to strict stationarity, it requires the existence of these moments.

**Definition 3.3.3.** The process \((X_t)\) is said to be **second-order stationary** if
- \( \mathbb{E}(X_t) = m \), for all \( t \in \mathbb{Z} \),
- \( \text{Cov}(X_t, X_{t+h}) = \gamma_X(h) \), for all \( h \in \mathbb{Z} \).

The function \( \gamma_X(\cdot) \left( \rho_X(\cdot) \equiv \frac{\gamma_X(\cdot)}{\gamma_X(0)} \right) \) is called the autocovariance function (autocorrelation function).

The simplest example of a second-order stationary process is white noise. This process is particularly important because it allows more complex stationary processes to be constructed.

**Definition 3.3.4.** The process \((\epsilon_t)\) is called **weak white noise**, if for some constant \( \sigma^2 > 0 \):
- \( \mathbb{E}(\epsilon_t) = 0 \), for all \( t \in \mathbb{Z} \),
CHAPTER 3. GAUSSIAN VERSUS NON GAUSSIAN FACTOR MODELS

- $\mathbb{E}(\epsilon_t^2) = \sigma^2$, for all $t \in \mathbb{Z}$,
- $\text{Cov}(\epsilon_t, \epsilon_{t+h}) = 0$, for all $t, h \in \mathbb{Z}, h \neq 0$.

Remark. Note that no independence assumption was made in the definition of weak white noise. The variables at different times are only uncorrelated and this distinction is particularly crucial for time series, however sometimes it is necessary to replace this assumption for a stronger one: the variables $\epsilon_t$ and $\epsilon_{t+h}$ are independent and identically distributed for all $t, h \in \mathbb{Z}$; the process is then said to be strong white noise.

The aim of time series analysis is to construct a model for the underlying stochastic process. This model is then used for analyzing the causal structure of the process and to obtain optimal predictors. The autoregressive moving average (ARMA) is the most widely used model for the prediction of second-order stationary processes. ARMA models can be viewed as a natural consequence of a fundamental result from 1938 due to Herman Wold.

**Theorem 35** (Wold’s Theorem). Any centered second-order stationary, and purely non-deterministic process, admits an infinite moving average representation of the form

$$X_t = \epsilon_t + \sum_{i=1}^{\infty} c_i \epsilon_{t-i},$$

(3.18)

where $(\epsilon_t)$ is the linear innovation process of $(X_t)$, that is

$$\epsilon_t = X_t - \mathbb{E}(X_t|\mathcal{H}_X(t-1)),$$

(3.19)

where $\mathcal{H}_X(t-1)$ denotes the Hilbert space generated by the random variables $X_{t-1}, X_{t-2}, \cdots$ and $\mathbb{E}(X_t|\mathcal{H}_X(t-1))$ denotes the orthogonal projection of $X_t$ onto $\mathcal{H}_X(t-1)$. The sequence of coefficients $(c_i)$ is such that $\sum_{i} c_i^2 < \infty$.


Remark. A stationary process $(X_t)$ is said to be purely deterministic if and only if $\bigcap_{n=-\infty}^{\infty} \mathcal{H}_X(n) = \{0\}$, where $\mathcal{H}_X(n)$ denotes, in the Hilbert space of the real, centered, and square integrable variables, the subspace constituted by the limits of the linear combinations of the variables $X_{n-i}, i \geq 0$. Therefore, for a purely non-deterministic (or regular) process, the linear past, sufficiently far away in the past, is of no use in predicting future values. See Brockwell, P. and Davis, R. (1991).

Remark. Note that in equation (3.19), the equivalence class $\mathbb{E}(X_t|\mathcal{H}_X(t-1))$ is identified with a random variable.

Truncating the infinite sum in (3.18), we obtain the process

$$X_t(q) = \epsilon_t + \sum_{i=1}^{q} c_i \epsilon_{t-i},$$

called a moving average process of order $q$, or MA$(q)$. We have that

$$||X_t - X_t(q)||^2_2 = \mathbb{E}\epsilon_t^2 \sum_{i>q} \epsilon_i^2 \to 0, \quad \text{as } q \to \infty.$$ 

It follows that the set of all finite order moving average processes is dense in the set of second order stationary and purely deterministic processes. The class of ARMA models is often preferred to the MA models for parsimony reasons, because they generally require fewer parameters.
Remark. A linear innovation process can be characterized by two properties i.e. \((\epsilon_t)\) is linear innovation process of \((X_t)\) if \((\epsilon_t)\) is white noise and \(\text{Cov}(\epsilon_t, X_{t-k}) = 0\) for all \(k > 0\).

In an ARMA model the current value \(X_t\) of the process is expressed as a linear function of the past values of the process and of current and past values of a white noise process \((\epsilon_t)\), i.e., of a sequence of uncorrelated and homoscedatic variables with mean zero.

**Definition 3.3.5.** A second-order stationary process \((X_t)\) is called ARMA\((p,q)\), where \(p\) and \(q\) are integers, if there exist real coefficients \(c, a_1, \cdots, a_p, b_1, \cdots, b_q\), such that

\[
X_t - \sum_{i=1}^{p} a_i X_{t-i} = c + \epsilon_t + \sum_{j=1}^{q} b_j \epsilon_{t-j}, \quad \text{for all } t \in \mathbb{Z},
\]

where \((\epsilon_t)\) is the linear innovation process of \((X_t)\).

**Theorem 36** (Characterization of an ARMA process). Let \((X_t)\) denote a second order stationary process. We have that

\[
\rho_X(h) + \sum_{i=1}^{p} a_i \rho_X(h-i) = 0, \quad \text{for all } |h| > q
\]

if and only if \((X_t)\) is an ARMA\((p,q)\) process.

**Proof.** See Brockwell, P. and Davis, R. (1991), pp. 89-90. \(
\)

Remark. For statistical convenience, ARMA models are often used under stronger assumptions on the noise than that of white noise. If in Definition 3.3.5, \((\epsilon_t)\) is assumed to be strong white noise, then \((X_t)\) is said to be a strong ARMA process.

Equation (3.20) can be rewritten symbolically in a more compact form:

\[
a(L)X_t = c + b(L)\epsilon_t,
\]

where \(a(z) = 1 - a_1 z - \cdots - a_p z^p\) and \(b(z) = 1 + b_1 z + \cdots + b_q z^q\) and \(L\) is the backward shift operator defined by \(L^j z_t = z_{t-j}\). The polynomials \(a(\cdot)\) and \(b(\cdot)\) are called autoregressive and moving average polynomials. The coefficients \(a_1, \cdots, a_p, b_1, \cdots, b_q\) are usually subject to stability constraints. These restrictions concern the zeros of the autoregressive and moving average polynomials:

\[
a(z) \equiv 1 - a_1 z - \cdots - a_p z^p \neq 0, \quad \text{for all } |z| \leq 1;
\]

\[
b(z) \equiv 1 + b_1 z + \cdots + b_q z^q \neq 0, \quad \text{for all } |z| \leq 1;
\]

Under these stability assumptions the polynomials \(a(\cdot)\) and \(b(\cdot)\) can be inverted, yielding two alternative representations of the process, an infinite moving average representation:

\[
X_t = a(L)^{-1}c + a(L)^{-1}b(L)\epsilon_t
\]

and an infinite autoregressive representation:

\[
b(L)^{-1}a(L)X_t = b(L)^{-1}c + \epsilon_t.
\]

The success of ARMA models is due to the tractability of the model, which is linear in the variables and some of the parameters at the same time. This first linearity feature yields forecast formulas that are easy to implement,
and the second one allows for estimation of the parameters using linear least squares methods. Note that making \( \mu = \frac{\sum a_i}{1 - \sum a_i} \) we can rewrite equation (3.20) as

\[
(X_t - \mu) - \sum_{i=1}^{p} a_i (X_{t-i} - \mu) = \epsilon_t + \sum_{j=1}^{q} b_j \epsilon_{t-j}.
\] (3.22)

Different methods have been proposed for the estimation of the parameters of ARMA models. Let \( (X_t) \) be an ARMA\((p,q)\) process, from equation (3.22) we have that

\[
(X_t - \mu) - a_1 (X_{t-1} - \mu) - \cdots - a_p (X_{t-p} - \mu) = \epsilon_t + b_1 \epsilon_{t-1} + \cdots + b_q \epsilon_{t-q}
\]

for some real constants \( \mu, a_1, \ldots, a_p, b_1, \ldots, b_q \). Assume that \( (\epsilon_t) \) is a sequence of i.i.d. random variables such that \( \epsilon_t \sim N(0, \sigma^2) \) for all \( t \in \mathbb{Z} \).

The parameters of this model can be estimated using a two step least squares method: first, compute \( \hat{\mu} \) using the sample average i.e. \( \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_t \); second, define \( Y_t = X_t - \hat{\mu} \) and regress \( Y_t \) on its \( p \) lags \( Y_{t-1}, \ldots, Y_{t-p} \) using ordinary least squares; then, for each of the \( q \) lags of the error take

\[
\hat{\epsilon}_t = Y_t - \sum_{i=1}^{p} \hat{\pi}_i Y_{t-i},
\]

where \( (\hat{\pi}_i)_{i=1}^{p} \) are the OLS estimates obtained in the previous step, write \( Y_t \) in the form

\[
Y_t = a_1 Y_{t-1} + \cdots + a_p Y_{t-p} + b_1 \hat{\epsilon}_{t-1} + \cdots + b_q \hat{\epsilon}_{t-q} + \epsilon_t
\]

and regress \( Y_t \) on \( Y_{t-1}, \cdots, Y_{t-p}, \hat{\epsilon}_{t-1}, \cdots, \hat{\epsilon}_{t-q} \) using ordinary least squares, to obtain estimates \( \hat{a}_1, \ldots, \hat{a}_p, \hat{b}_1, \ldots, \hat{b}_q \) of the coefficients. See Durbin, J. (1960).

Maximum likelihood is widely used in the estimation of ARMA models. The estimation of the parameters can be done in two different ways. One approach is to try to determine the joint density \( f(X_1, \ldots, X_n; \theta) \) directly which requires among other things estimates of the \( n \times n \) covariance matrix. An alternative approach relies on the factorization of the joint density into a series of conditional densities and the density of a set of initial values:

\[
L(X^n, \theta) = \sum_{t=p+1}^{n} \log f(X_t | X_{t-1}; \theta) + \log f(X_p, \theta).
\]

(3.23)

Note that \( X^n \) is a \( k \)-dimensional vector with components \( X_1, X_2, \ldots, X_k \) and \( \theta = (\mu, a_1, \ldots, a_p, b_1, \ldots, b_q) \) denotes the vector of parameters from equation (3.22). Two types of maximum likelihood estimates can be computed. The first type is obtained maximizing the conditional log-likelihood function (first term in equation (3.23)), while the second type requires maximizing the exact log-likelihood function \( L(X^n, \theta) \). For stationary models, both estimators are consistent and have the same limiting distribution.

### 3.4 Financial series and GARCH models

Modelling financial series is a complex problem. This complexity is mainly due to the existence of statistical regularities (stylized facts) which are common to a large number of financial series and are difficult to reproduce.
artificially using stochastic models. Most of these stylized facts were put forward in a paper by Mandelbrot, B. (1963). Since then they have been documented and completed by many empirical studies. Let \( r_t = \frac{p_t - p_{t-1}}{p_{t-1}} \) be a series of daily price returns. The following properties have been amply commented upon in the financial literature:

- Sample paths of returns are generally compatible with the second order stationarity assumption. The returns oscillate around zero and while the oscillations vary a great deal in magnitude they are almost constant in average over long subperiods.
- The series of price returns generally displays small autocorrelations, making it close to white noise.
- Squares returns \( (r_t^2) \) or absolute returns \( (|r_t|) \) are generally strongly autocorrelated.
- Volatility clustering. Large absolute returns \( |r_t| \) tend to appear in clusters. Turbulent (high-volatility) subperiods are followed by quiet (low-volatility) periods.
- Fat tails. When the empirical distribution of daily returns is drawn, one can generally observe that it does not resemble a Gaussian distribution. The densities have fat tails and are sharply peaked at zero: they are called leptokurtic. A measure of the leptokurticity is the kurtosis coefficient, defined as the ratio of the sample fourth order moment to the squared sample variance. It is asymptotically equal to 3 for Gaussian i.i.d. distributions. This coefficient is much greater than 3 for return series.
- Leverage Effect. The so called leverage effect was noted by Black, F. (1976), and involves asymmetry of the impact of past positive and negative values on the current volatility. Negative returns (corresponding to price decreases) tend to increase volatility by a larger amount than positive returns (price increases) of the same magnitude.

Any satisfactory model for daily returns must be able to capture the main stylized facts described in the previous section. Particularly important are, the leptokurticity, the unpredictability of returns, and the existence of positive autocorrelations in the square and absolute value returns.

The fact that large absolute returns tend to be followed by large absolute returns (whatever the sign of the price variations might be) is hardly compatible with the assumption of constant conditional variance. This phenomenon is called conditional heteroscedasticity:

\[
Var(r_t|r_{t-1}, r_{t-2}, \cdots) \neq \text{constant}.
\]

Remark. Note that conditional heteroscedasticity is perfectly compatible with stationarity (in the strict and second order senses).

It is time now to introduce the concept of information set. The information set at time \( t \) is denoted by \( \Omega_t \) and contains all events measurable at time \( t \), in particular it contains realizations of all variables that have occurred on or before \( t \). Given a time series \( (X_t) \), we denote the conditional expectation and conditional variance of the stochastic variable \( X_t \) given the information set at time \( s \) by \( \mathbb{E}(X_t|\Omega_s) \) and \( \text{Var}(X_t|\Omega_s) \) respectively.

**Proposition 37.** If \( (\epsilon_t) \) is a strong ARMA(\( p,q \)) process i.e. there exist constants \( c, a_1, \cdots, a_p, b_1, \cdots, b_q \) such that

\[
\epsilon_t = c + \sum_{i=1}^{p} a_i \epsilon_{t-i} + z_t + \sum_{j=1}^{q} b_j z_{t-j},
\]

where \( (z_t) \) is strong white noise with mean 0 and variance \( \sigma_z^2 \). Then the variance of \( \epsilon_t \) conditional on the information set at time \( t-1 \) is constant and equal to \( \sigma_z^2 \).
Proof. Let \((\epsilon_t)\) be a (strong) ARMA\((p,q)\) process, then

\[
E(\epsilon_t | \Omega_{t-1}) = c + \sum_{i=1}^{p} a_i \epsilon_{t-i} + \sum_{j=1}^{q} b_j z_{t-j}.
\]

Therefore,

\[
\epsilon_t - E(\epsilon_t | \Omega_{t-1}) = z_t.
\]

Hence,

\[
Var(\epsilon_t | \Omega_{t-1}) = E[(\epsilon_t - E(\epsilon_t | \Omega_{t-1}))^2 | \Omega_{t-1}] = E[z_t^2 | \Omega_{t-1}] = E[z_t^2] = \sigma_z^2.
\]

Suppose that we have noticed that recent daily returns have been unusually volatile. We might expect that tomorrow’s return will also be more variable than usual, however an ARMA model cannot capture this type of behaviour because its conditional variance given the past is constant. Thus, other models must be considered if we want to account for non-constant volatility.

Autoregressive conditionally heteroscedastic (ARCH) models were introduced by Engle, R. (1982) and their GARCH (generalized ARCH) extension is due to Bollerslev, T. (1986). In these models the key concept is the conditional variance, that is the variance conditional on the past. In the classical ARCH models the variance is expressed as a linear function of squared past values of the series. This specification is able to capture the main stylized facts characterizing financial series.

**Definition 3.4.1.** A process \((\epsilon_t)\) is a (strong) ARCH\((q)\) process if there exist \(q \in \mathbb{Z}\) and real constants \(\omega > 0, \alpha_i \geq 0\) for \(i = 1, \cdots, q\) such that

\[
\epsilon_t = \sqrt{h_t} z_t
\]

\[
h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2,
\]

where \((z_t)\) is a sequence of i.i.d. random variables with \(E(z_t) = 0\) and \(Var(z_t) = 1\) for all \(t \in \mathbb{Z}\).

**Theorem 38** (Properties of an ARCH\((q)\) process). Let \((\epsilon_t)\) be a second order stationary ARCH\((q)\) process i.e.

\[
\epsilon_t = \sqrt{h_t} z_t
\]

\[
h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2,
\]

with \(\omega > 0\) and \(\alpha_i \geq 0\) for \(i = 1, \cdots, q\), then

\[
E(\epsilon_t | \Omega_{t-1}) = 0
\]

\[
Var(\epsilon_t | \Omega_{t-1}) = w + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1}^2
\]

\[
E(\epsilon_t) = 0
\]

\[
Var(\epsilon_t) = \frac{w}{1 - \sum_{i=1}^{q} \alpha_i},
\]

where \(E(\epsilon_t | \Omega_{t-1})\) and \(Var(\epsilon_t | \Omega_{t-1})\) are the conditional expectation and conditional variance of \(\epsilon_t\) given the information set at time \(t - 1\), while \(E(\epsilon_t)\) and \(Var(\epsilon_t)\) are the unconditional expectation and variance of the process \((\epsilon_t)\).
Proof. We have that

\[
E[\epsilon_t | \Omega_{t-1}] = E \left[ \left( \sqrt{\omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2} \right) z_t | \Omega_{t-1} \right]
\]

\[
= \left( \sqrt{\omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2} \right) E[z_t | \Omega_{t-1}]
\]

\[
= 0,
\]

and

\[
Var(\epsilon_t | \Omega_{t-1}) = E \left[ \left( \left( \sqrt{\omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2} \right) z_t - 0 \right)^2 | \Omega_{t-1} \right]
\]

\[
= E \left[ \left( \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \right) z_t^2 | \Omega_{t-1} \right]
\]

\[
= (\omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2) E[z_t^2 | \Omega_{t-1}]
\]

\[
= \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2.
\]

In addition,

\[
E [\epsilon_t] = E \left[ \sqrt{\omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2} \right] E[z_t]
\]

\[
= 0,
\]

and

\[
Var(\epsilon_t) = E[\epsilon_t^2] - (E[\epsilon_t])^2
\]

\[
= E[\epsilon_t^2]
\]

\[
= E[h_t z_t^2]
\]

\[
= E[h_t]
\]

\[
= \omega + \sum_{i=1}^{q} \alpha_i E[\epsilon_{t-i}^2]
\]

\[
= \omega + \sum_{i=1}^{q} \alpha_i Var(\epsilon_{t-i}).
\]

Because the process \((\epsilon_t)\) is assumed to be second order stationary, we have that

\[
Var(\epsilon_t) = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i}.
\]

\[\square\]

Corollary 38.1. If \((\epsilon_t)\) is a second order stationary ARCH \((q)\) process then \(\sum_{i=1}^{q} \alpha_i < 1\).

Definition 3.4.2. A process \((\epsilon_t)\) is called (strong) GARCH \((p, q)\) if there exist integers \(p\) and \(q\) and real coefficients \(\omega, \alpha_i\) and \(\beta_j\) such that for all \(t \in \mathbb{Z}\)

\[
\epsilon_t = \sqrt{h_t} z_t,
\]

\[
h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j},
\]

(3.26)

where \((z_t)\) is a sequence of i.i.d. random variables with \(E(z_t) = 0\) and \(Var(z_t) = 1\) for all \(t \in \mathbb{Z}\) and \(\omega > 0\), \(\alpha_i \geq 0\) and \(\beta_j \geq 0\) for all \(i = 1, \ldots, q\) and \(j = 1, \ldots, p\) respectively.

Definition 3.4.3. Let \((\epsilon_t)\) be a GARCH \((p, q)\) process, if the coefficients \(\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p\) of the process satisfy that

\[
\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = 1,
\]

then \((\epsilon_t)\) is called an integrated GARCH \((p, q)\) or IGARCH \((p, q)\) process.
Theorem 39 (Properties of a GARCH(p, q) process). Let \((\epsilon_t)\) be a second order stationary GARCH\((p, q)\) process:

\[
\epsilon_t = \sqrt{h_t} z_t \\
h_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j},
\]

with \(p, q \in \mathbb{Z}, \omega > 0, \alpha_i \geq 0\) and \(\beta_j \geq 0\) for all \(i = 1, \ldots, q\) and \(j = 1, \ldots, p\) respectively, then

\[
\begin{align*}
E(\epsilon_t | \Omega_{t-1}) & = 0 \\
Var(\epsilon_t | \Omega_{t-1}) & = w + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \\
E(\epsilon_t) & = 0 \\
Var(\epsilon_t) & = \frac{w}{1 - (\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j)},
\end{align*}
\]

(3.27)

where \(E(\epsilon_t | \Omega_{t-1})\) and \(Var(\epsilon_t | \Omega_{t-1})\) are the conditional expectation and conditional variance of \(\epsilon_t\) given the information set at time \(t - 1\), while \(E(\epsilon_t)\) and \(Var(\epsilon_t)\) are the unconditional expectation and variance of the process \((\epsilon_t)\).

Proof. The proofs of the first three equations of (3.27) are almost identical to the proofs of the first three equations of (3.25), which were already provided, therefore we only show the derivation of the last equation.

\[
\begin{align*}
Var(\epsilon_t) & = E[\epsilon_t^2] - (E[\epsilon_t])^2 \\
& = E[\epsilon_t^2] \\
& = E[h_t z_t^2] \\
& = E[h_t] \\
& = \omega + \sum_{i=1}^{q} \alpha_i E[\epsilon_{t-i}^2] + \sum_{j=1}^{p} \beta_j E[h_{t-j}] \\
& = \omega + \sum_{i=1}^{q} \alpha_i Var(\epsilon_{t-i}) + \sum_{j=1}^{p} \beta_j E[h_{t-j}].
\end{align*}
\]

However,

\[
\begin{align*}
E[h_{t-j}] & = E[h_{t-j} z_{t-j}^2] \\
& = E[\epsilon_{t-j}^2] \\
& = Var(\epsilon_{t-j}).
\end{align*}
\]

Therefore,

\[
Var(\epsilon_t) = \omega + \sum_{i=1}^{q} \alpha_i Var(\epsilon_{t-i}) + \sum_{i=1}^{p} \beta_j Var(\epsilon_{t-j}).
\]

Because the process \((\epsilon_t)\) is assumed to be second order stationary, we have that

\[
Var(\epsilon_t) = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}.
\]

\(\square\)

Remark. Note that for all \(h, t \in \mathbb{Z}, h \neq t:\)

\[
Cov(\epsilon_t, \epsilon_{t+h}) = E(\epsilon_t \epsilon_{t+h}) - E(\epsilon_t)E(\epsilon_{t+h}) = 0.
\]

Corollary 39.1. If \((\epsilon_t)\) is a second order stationary GARCH\((p, q)\) process then \(\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1\).

Let \((\epsilon_t)\) be a GARCH\((p, q)\) process and define \(u_t = \epsilon_t^2 - h_t\). Substituting \(h_{t-j} = \epsilon_{t-j}^2 - u_{t-j}\) in (3.26) the
following representation is obtained:

\[
\epsilon_t^2 = \omega + \sum_{i=1}^{r} (\alpha_i + \beta_i) \epsilon_{t-i}^2 + u_t - \sum_{j=1}^{p} \beta_j u_{t-j}, \quad t \in \mathbb{Z},
\]

(3.28)

where \( r = \max(p, q) \), with the convention \( \alpha_i = 0 \) (\( \beta_i = 0 \)) if \( i > q \) (\( j > p \)).

Note that \( u_t = \epsilon_t^2 - h_t = h_t(z_t^2 - 1) \). Therefore \( \mathbb{E}(u_t) = \mathbb{E}(z_t^2 - 1)\mathbb{E}(h_t) = 0 \), for all \( t \in \mathbb{Z} \) because \( (z_t) \) is assumed to be i.i.d with zero mean and variance 1. Also, for \( k > 0 \) and \( t \in \mathbb{Z} \), we have that

\[
\text{Cov}(u_t, u_{t-k}) = \mathbb{E}(h_t(z_t^2 - 1)h_{t-k}(z_{t-k}^2 - 1)) = \mathbb{E}(z_t^2 - 1)\mathbb{E}(h_th_{t-k}(z_{t-k}^2 - 1)) = 0,
\]

\[
\text{Cov}(u_t, \epsilon_{t-k}^2) = \mathbb{E}(h_t(z_t^2 - 1)\epsilon_{t-k}^2) = \mathbb{E}(z_t^2 - 1)\mathbb{E}(h_t\epsilon_{t-k}^2) = 0.
\]

Thus, equation (3.28) has the structure of an ARMA model. In fact, under additional assumptions (implying the second-order stationarity of \( (\epsilon_t^2) \)), if \( (\epsilon_t) \) is a GARCH\((p,q)\) process, then \( (\epsilon_t^2) \) is an ARMA\((r,p)\) process.

Since the introduction of ARCH models by Engle in 1982 different methods have been proposed for the estimation of the parameters of both ARCH and GARCH models. Some of the most popular ones are described below.

The simplest estimation method for ARCH models is that of ordinary least squares (OLS). Consider the ARCH\((q)\) model:

\[
\epsilon_t = \sqrt{h_t} z_t,
\]

\[
h_t = \omega_0 + \sum_{i=1}^{q} \alpha_0 i \epsilon_{t-i}^2, \text{ with } \omega_0 > 0, \alpha_0_i \geq 0, i = 1, \cdots, q,
\]

\((z_t)\) sequence of i.i.d. random variables, with \( \mathbb{E}(z_t) = 0, \text{Var}(z_t) = 1 \).

The OLS method (proposed by Engle in 1982) uses the autoregressive representation on the squares of the observed process. No assumption is made on the law of \( (z_t) \). From (3.28), taking \( u_t = \epsilon_t^2 - h_t \), the following AR\((q)\) representation of \( (\epsilon_t) \) is obtained:

\[
\epsilon_t^2 = \omega_0 + \sum_{i=1}^{q} \alpha_0 i \epsilon_{t-i}^2 + u_t.
\]

(3.30)

Let \( \theta_0 \) denote the true value of the vector of parameters i.e. \( \theta_0 = (\omega_0, \alpha_{01}, \cdots, \alpha_{0q})' \) and \( \theta \) be a generic vector of parameters. Assume that \( \epsilon_1, \cdots, \epsilon_n \) is a realization of length \( n \) of the process \( (\epsilon_t) \), let \( \epsilon_0, \cdots, \epsilon_{1-q} \) be initial values (which can be chosen, for example, equal to zero) and define \( V_{t-1}' = (1, \epsilon_{t-1}', \cdots, \epsilon_{t-q}') \).

From (3.30) we have that

\[
\epsilon_t^2 = V_{t-1}' \theta_0 + u_t, \quad t = 1, \cdots, n,
\]

which can be rewritten as \( Y = X \theta_0 + U \), with

\[
Y = \begin{bmatrix}
\epsilon_1^2 \\
\vdots \\
\epsilon_n^2
\end{bmatrix}, \quad X = \begin{bmatrix}
1 & \epsilon_0^2 & \cdots & \epsilon_{q+1}^2 \\
\vdots & \ddots & \ddots & \vdots \\
1 & \epsilon_{n-1}^2 & \cdots & \epsilon_{n-q}^2
\end{bmatrix}, \quad U = \begin{bmatrix}
u_1 \\
\vdots \\
u_n
\end{bmatrix},
\]

Therefore, assuming that the matrix \( X'X \) is invertible, the OLS estimator of \( \theta_0 \) is given by

\[
\hat{\theta} = \arg \min_{\theta} ||Y - X \theta||^2 = (X'X)^{-1} X'Y.
\]

(3.31)

Note that the estimated parameters from (3.31) can be negative, which is problematic as predictions of the volatil-
ity can be negative as well. In order to avoid this problem the constrained OLS estimator is used instead:

$$\hat{\theta}^c = \arg \min_{\theta \in [0, \infty)^{q+1}} ||Y - X\theta||^2.$$ 

Note that in (3.30) the errors \((u_t)\) satisfy that \(u_t = \epsilon_t^2 - h_t = (z_t^2 - 1)h_t\), therefore the process \((u_t)\) is a conditionally heteroscedastic process, with conditional variance:

$$\text{Var}(u_t | \Omega_{t-1}) = (\kappa_z - 1)h_t^2, \quad \kappa_z = E(z_t^4).$$

A feasible generalized least squares estimator of \(\theta_0\) can be defined as follows:

For all \(\theta = (\omega, \alpha_1, \cdots, \alpha_q)'\), let 

$$h_t(\theta) = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 \quad \text{and} \quad \hat{\Omega} = \text{diag}(h_1^{-2}(\hat{\theta}), \cdots, h_n^{-2}(\hat{\theta})).$$

The FGLS estimator is defined by

$$\hat{\theta}_F = (X'\hat{\Omega}X)^{-1}X'\hat{\Omega}Y.$$ 

This estimator was developed, by Bose, A. and Mukherjee, K. (2003).

Franq, C. and Zakoian, J. (2010) prove the consistency of both estimators under certain regularity assumptions, and compare the asymptotic properties of both methods. An OLS estimator can be defined for a GARCH process too, however, the estimator is not explicit because \((\epsilon_t)\) is not an AR process. However, the most widely used method to estimate the parameters of a strong GARCH processes is (Gaussian) quasi-maximum likelihood. The QML estimator for ARCH models was first proposed by Engle, R. (1982), and extended for GARCH models by Bollerslev, T. (1986).

Suppose that the observations \(\epsilon_1, \cdots, \epsilon_n\) constitute a realization of a strong GARCH \((p, q)\) process i.e.

$$\epsilon_t = \sqrt{h_t}z_t,$$

$$h_t = \omega_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}, \quad \text{for all } t \in \mathbb{Z},$$

where \((z_t)\) is a sequence of i.i.d. variables with mean zero and variance 1, \(\omega_0 > 0, \alpha_i \geq 0 (i = 1, \cdots, q)\) and \(\beta_j \geq 0, (j = 1, \cdots, p)\). The orders of \(p and q\) are assumed to be known. The vector of parameters

$$\theta = (\theta_1, \cdots, \theta_{p+q+1})' = (\omega, \alpha_1, \cdots, \alpha_q, \beta_1, \cdots, \beta_p)'$$

belongs to a parameter space of the form

$$\Theta \subset (0, +\infty) \times [0, \infty)^{p+q}.$$ 

The true value of the parameter is unknown, and is denoted by \(\theta_0 = (\omega_0, \alpha_{01}, \cdots, \alpha_{0q}, \beta_1, \cdots, \beta_{0p})'\). Given initial values \(\epsilon_0, \cdots, \epsilon_{n-q}, h_0, \cdots, h_{1-p}\), the conditional Gaussian quasi-likelihood is given by

$$L_n(\theta) = L_n(\theta; \epsilon_1, \cdots, \epsilon_n) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi h_t}} \exp \left( -\frac{\epsilon_t^2}{2h_t} \right),$$

(3.32)
where the $\tilde{h}_t$ are recursively defined, for $t \geq 1$, by

$$\tilde{h}_t = \tilde{h}_t(\theta) = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_j \tilde{h}_{t-j}.$$ 

For a given $\theta$, a reasonable set of initial values is

$$\epsilon_0^2 = \cdots = \epsilon_{1-q}^2 = h_0 = \cdots = h_{1-p} = \omega$$

where

$$\tilde{h}_t = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j}.$$ 

However, this set of initial values is not suitable when $(\epsilon_t)$ is not second order stationary, which can happen if $(\epsilon_t)$ is an IGARCH process for example. Then, more suitable sets of initial values are

$$\epsilon_0^2 = \cdots = \epsilon_{1-q}^2 = h_0 = \cdots = h_{1-p} = \omega,$$

or

$$\epsilon_0^2 = \cdots = \epsilon_{1-q}^2 = h_0 = \cdots = h_{1-p} = \epsilon_1^2.$$ 

A QML estimator of $\theta$ can be obtained maximizing (3.32) with respect to $\theta$ i.e.

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L_n(\theta)$$

or, minimizing $L_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left( \epsilon_t^2 + \log \tilde{h}_t \right)$, i.e.

$$\hat{\theta} = \arg\min_{\theta \in \Theta} L_n(\theta).$$

Franq, C. and Zakoian, J. (2010) show that the QML estimator is consistent and asymptotically normal under certain regularity conditions and that the choice of initial values is unimportant for the asymptotic properties of the QML estimator.

### 3.5 Model approaches

A popular method to estimate time-varying alphas and betas in a dynamic 1-factor model is via rolling window regression. Assuming that each window starts at $t - \tau$ and ends at $t - 1$, our models satisfy the following equation for each fixed $t \geq \tau + 1$:

$$R_{f,t-j} = \alpha_t + \beta_t R_{m,t-j} + \epsilon_{t-j}, \quad \text{for all } j = 1, \cdots, \tau. \tag{3.33}$$

$R_{f,t-j} = r_{f,t-j} - r_{F,t-j}$ and $R_{m,t-j} = r_{m,t-j} - r_{F,t-j}$ where $r_{f,t-j}$, $r_{m,t-j}$ and $r_{F,t-j}$ denote the daily return of an asset, the daily return of the market and the daily risk free rate respectively, at time $t - j$. The parameter $\beta_t$ denotes the exposure of the asset to the market at time $t$, while $\alpha_t$ denotes the excess return of the asset over the market excess return given the exposure of the asset to the market. In particular, if the asset is a hedge fund, then alpha is often considered an indicator of the hedge fund’s manager ability to generate positive returns.

We consider different versions of the model in (3.33). The first one assumes that the residuals $(\epsilon_{t-j})$ are i.i.d with $\epsilon_{t-j} \sim N(0, \sigma_{\epsilon,t}^2)$. The other three models incorporate dynamic volatility assumptions on different components
of (3.33).
For simplicity, let us rewrite equation (3.33) as
\[ R_{f,t} = \alpha_s + \beta_s R_{m,t} + \epsilon_t, \quad \text{for all } t = s - \tau, \ldots, s - 1, \quad (3.34) \]
where \( s \geq \tau + 1 \).

### 3.5.1 Model 1 (Gaussian white noise residuals)

The first model that we consider is a 1-factor model with time-varying parameters based on the classical CAPM equation introduced independently by Sharpe, W. (1964) and Lintner, J. (1965). For each \( s \geq \tau + 1 \) we have that:
\[ R_{f,t} = \alpha_s + \beta_s R_{m,t} + \epsilon_t, \quad (\epsilon_t) \text{ i.i.d } N(0, \sigma^2_{\epsilon,s}) \text{ distributed, for all } t = s - \tau, \ldots, s - 1. \quad (3.35) \]

The parameters of equation (3.35) can be estimated using either ordinary least squares or maximum likelihood. Both methods yield the same estimates. These procedures are described in detail in section 3.2.

### 3.5.2 Model 2 (GARCH residuals)

The second model assumes that the residuals \( (\epsilon_t) \) follow a GARCH(1,1) process. For \( s \geq \tau + 1 \), we have that:
\[ R_{f,t} = \alpha_s + \beta_s R_{m,t} + \epsilon_t, \quad (3.36) \]
\[ h_t = \omega_s + \alpha_{1,s} \epsilon_{t-1}^2 + \beta_{1,s} h_{t-1}, \text{ where } \alpha_{1,s}, \beta_{1,s} \geq 0 \text{ and } \omega_s > 0. \]

In general, a regression model with (strong) GARCH(\( p, q \)) errors can be written as:
\[ Y_t = X_t \beta + \epsilon_t, \quad (3.37) \]
\[ \epsilon_t = \sqrt{h_t} z_t, \text{ where } (z_t) \text{ i.i.d. with } \mathbb{E}(z_t) = 0 \text{ and } \text{Var}(z_t) = 1, \]
\[ h_t = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}, \]
where \( \alpha_i \) and \( \beta_j \) are nonnegative constants and \( \omega \) is a strictly positive constant, \( Y_t \) is the dependent variable, \( X_t \) is a vector of exogenous or predetermined regressors and \( \beta \) is a vector of parameters.

The regression model with ARCH errors was proposed by Engle, R. (1982) along with two different methods that yield consistent estimates of the parameters of the model. The properties of these estimators are described in detail in Gourieroux, C. (1997). The main estimation procedures for (3.37) are extensions to the GARCH case of the two methods proposed by Engel in 1982. See Davidson, R. and MacKinnon, J. (2004).

The first estimation method is a three step regression procedure. First, a consistent estimate of \( \beta \) is obtained using the ordinary least squares estimator from the regression of \( Y_t \) on \( X_t \), denoted by \( \tilde{\beta} \). From this estimation we obtain as residuals
\[ \tilde{\epsilon}_t = Y_t - X_t \tilde{\beta}. \]

The parameters \( \omega, \alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p \) appearing in the conditional variance can be estimated using any of the consistent estimators of GARCH models discussed in Section 3.4, the estimated parameters are denoted by \( \hat{\omega}, \hat{\alpha}_1, \ldots, \hat{\alpha}_q, \hat{\beta}_1, \ldots, \hat{\beta}_p \). Finally the estimator of \( \beta \) is improved applying weighted least squares to a regression
of \( Y \) on \( X \) using the conditional variances \( \hat{h}_t \) from the previous step as diagonal elements of the weighting matrix.

The second step estimator can be denoted by \( \tilde{\beta} \).

The most popular method to estimate the parameters of (3.37) is quasi-maximum likelihood.

Let \( \theta \) be the vector of parameters and consider the following conditional mean and conditional variance:

\[
\begin{align*}
    m_t(\theta) &= E(Y_t|Y_{t-1}, X_t) \\
    \sigma^2_t(\theta) &= Var(Y_t|Y_{t-1}, X_t).
\end{align*}
\]

These two conditional moments depend on past values of the process, and current and past values of some exogenous regressors \( X \). A conditional (Normal) quasi-likelihood function is obtained by assuming that the conditional distribution of \( Y_t \) given \( Y_{t-1}, X_t \) is normal. This quasi-likelihood function is

\[
L_n(\theta) = \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi \sigma^2_t(\theta)}} \exp\left(-\frac{(Y_t - m_t(\theta))^2}{2\sigma^2_t(\theta)}\right).
\]

Taking logarithm, we obtain the following conditional quasi-loglikelihood function:

\[
\log L_n(\theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} n \sum_{i=1}^{n} \log(\sigma^2_i(\theta)) - \frac{1}{2} \sum_{i=1}^{n} \frac{(Y_t - m_t(\theta))^2}{\sigma^2_i(\theta)}.
\]

(3.38)

Note that

\[
\begin{align*}
    m_t(\theta) &= X_t \beta, \\
    \sigma^2_t(\theta) &= h_t \\
    &= \omega + \sum_{i=1}^{q} \alpha_i (Y_{t-i} - X_{t-i} \beta)^2 + \sum_{j=1}^{p} \beta_j h_{t-j}.
\end{align*}
\]

(3.39)

A QML estimator of \( \theta \) can be obtained maximizing (3.38) with respect to \( \theta \) i.e. \( \hat{\theta} = \arg \max_{\theta} \log L_n(\theta) \).

We estimated the parameters of (3.36) using both the three step regression procedure an the conditional quasi-loglikelihood method.

Remark. Assuming that \( z_t \sim N(0,1) \) for all \( t \) (in addition to \( z_t \) being i.i.d.), implies that

\[
Y_t|Y_{t-1}, X_t \sim N(X_t \beta, h_t).
\]

Therefore, the assumption of \( z_t \) being i.i.d. and \( z_t \sim N(0,1) \) for all \( t \) is sometimes used instead of directly assuming that the conditional distribution of \( Y_t \) given \( Y_{t-1} \) and \( X_t \) is conditionally Gaussian.

### 3.5.3 Model 3 (Unobserved GARCH)

Our third model assumes that the centered regressor follows a GARCH(1,1) process.

For \( s \geq \tau + 1 \), we have that:

\[
\begin{align*}
    R_{f,t} &= \alpha_s + \beta_s R_{m,t} + \eta_t, \quad \text{for all } t = s - \tau, \cdots, s - 1. \\
    \epsilon_t &= \sqrt{h_t} z_t, \\
    h_t &= \omega_s + \alpha_{1,s} \epsilon_{t-1}^2 + \beta_{1,s} h_{t-1}, \quad \text{with } \alpha_{1,s}, \beta_{1,s} \geq 0 \text{ and } \omega_s > 0.
\end{align*}
\]

(3.40)

where \( R_{m,t} = \epsilon_t + \mu_{m,s} \) (note that \( \mu_{m,s} = \mathbb{E}(R_{m,t}) \)), \( (z_t) \) i.i.d. with \( z_t \sim N(0,1) \), \( (\eta_t) \) i.i.d. with \( \eta_t \sim N(0,\sigma^2_{\eta,s}) \) and the disturbances \( (z_t) \) and \( (\eta_t) \) are also assumed to be mutually independent. For a given fixed window, the model in (3.40) belongs to a class often called unobserved GARCH or GARCH with errors. This
family of models was first proposed by Harvey, Ruiz and Sentana in 1982. A number of papers have focused on these models. See Harvey, A., Ruiz, E. and Sentana, E. (1992), Gourieroux, C., Monfort, A. and Renault, E. (1993) and Franq, C. and Zakoian, J. (2000).

\[ R_{f,t} = \alpha_s + \beta_s (\epsilon_t + \mu_{m,s}) + \eta_t \] is conditionally Gaussian, with conditional mean and conditional variance given by:

\[
\begin{align*}
\mathbb{E}(R_{f,t} | \Omega_{t-1}) &= \alpha_s + \beta_s \mu_{m,s}, \\
\text{Var}(R_{f,t} | \Omega_{t-1}) &= \beta^2_h + \sigma^2_{\eta,s}.
\end{align*}
\]

respectively.

Let \( \theta_s \) be the vector of parameters of (3.40) i.e. \( \theta_s = (\alpha_s, \beta_s, \sigma^2_{\eta,s}, \omega_s, \alpha_{s,1}, \beta_{s,1})' \), a conditional quasi-likelihood function is:

\[
L_r(\theta_s) = \prod_{t=s-\tau}^{s-1} \frac{1}{\sqrt{2\pi(\beta^2_s h_t + \sigma^2_{\eta,s})}} \exp \left( -\frac{(R_{f,t} - (\alpha_s + \beta_s \mu_{m,s}))^2}{2(\beta^2_s h_t + \sigma^2_{\eta,s})} \right).
\]

Therefore, a conditional quasi-loglikelihood function is given by:

\[
\log L_r(\theta_s) = -\frac{\tau}{2} \log(2\pi) - \frac{1}{2} \sum_{t=s-\tau}^{s-1} \log(\beta^2_s h_t + \sigma^2_{\eta,s}) - \frac{1}{2} \sum_{t=s-\tau}^{s-1} \frac{(R_{f,t} - (\alpha_s + \beta_s \mu_{m,s}))^2}{\beta^2_s h_t + \sigma^2_{\eta,s}}. \tag{3.41}
\]

An estimator of \( \theta_s \) can be obtained maximizing (3.41) i.e. \( \hat{\theta}_s = \arg\max_{\theta_s \in \Theta_s} \log L_r(\theta_s) \).

The following procedure is used to estimate the parameters in (3.40). First, the sample mean of \( \{R_{f,m,t} \} \) is computed, i.e. \( \bar{\mu}_{m,s} = \frac{1}{\tau} \sum_{t=s-\tau}^{s-1} R_{f,m,t} \). Define \( \tilde{\epsilon}_t = R_{f,m,t} - \bar{\mu}_{m,s} \), because \( \{\epsilon_t\} \) is assumed to follow a GARCH(1,1) process we can use conditional quasi-likelihood to estimate \( \omega_s, \alpha_{s,1} \) and \( \beta_{s,1} \), see Section 3.4. The estimated parameters are denoted by \( \hat{\omega}_s, \hat{\alpha}_{s,1}, \hat{\beta}_{s,1} \). These estimates are substituted in equation (3.41), we obtain that

\[
\hat{L}_r(\alpha_s, \beta_s, \sigma^2_{\eta,s}) = -\frac{\tau}{2} \log(2\pi) - \frac{1}{2} \sum_{t=s-\tau}^{s-1} \log(\beta^2_s \tilde{h}_t + \sigma^2_{\eta,s}) - \frac{1}{2} \sum_{t=s-\tau}^{s-1} \frac{(R_{f,t} - (\alpha_s + \beta_s \hat{\mu}_{m,s}))^2}{\beta^2_s \tilde{h}_t + \sigma^2_{\eta,s}}. \tag{3.42}
\]

where \( \tilde{h}_t = \hat{\omega}_s + \hat{\alpha}_{1,s} \tilde{\epsilon}_{t-1}^2 + \hat{\beta}_{1,s} \tilde{h}_{t-1} \). Finally (3.42) is maximized with respect to \( \alpha_s, \beta_s \) and \( \sigma^2_{\eta,s} \). This last step yields the estimates of \( \alpha_s, \beta_s \) and \( \sigma^2_{\eta,s} \).

### 3.5.4 Model 4 (Weak GARCH residuals)

Before our last model is described and a procedure to estimate the model parameters is proposed, we must introduce some concepts and results. The notions of temporal aggregation, contemporaneous aggregation, weak GARCH and weak ARMA-GARCH models are particularly relevant.

The temporal aggregation problem can be formulated as follows: given a process \( \{X_t\} \) and an integer \( m \), what are the properties of the sampled process \( \{X_{mt}\} \) (that is constructed from \( \{X_t\} \) by only keeping the \( m \)-th observation)? Does the aggregated process \( \{X_{mt}\} \) belong to the same class of models as the original \( \{X_t\} \)? If this holds for any \( X_t \) and \( m \), the class is said to be stable by temporal aggregation.

The class of weak GARCH models was first introduced by Drost, F. and Nijman, T. (1993) while studying the temporal aggregation problem for GARCH models. They showed that although the class of GARCH models is, in general, not closed under temporal aggregation, there exists a wider class of GARCH type models that is stable under temporal aggregation, the class of weak GARCH models.

It is standard in finance to consider linear combinations of several series (for instance, portfolios), the contem-
For a fourth order stationary process \( (\epsilon_t) \) is said to be weak GARCH \((p, r)\) if:

1. \( (\epsilon_t) \) is a white noise,
2. \( (\epsilon_t^2) \) admits an ARMA representation of the form

\[
e_t^2 - \sum_{i=1}^{r} a_i \epsilon_{t-i}^2 = c + u_t + \sum_{j=1}^{p} b_j u_{t-j},
\]

where \( (u_t) \) is the linear innovation process of \( (\epsilon_t^2) \).

**Remark.** Recall that if \( (u_t) \) is the linear innovation process of \( (\epsilon_t^2) \) then \( (u_t) \) is weak white noise and for all \( k > 0 \)
\[
\text{Cov}(u_t, \epsilon_{t-k}^2) = 0.
\]

**Remark.** Definition 3.5.1 (which can be found in Franq, C. and Zakoian, J. (2010)) is more general than that of weak GARCH proposed by Drost, F. and Nijman, T. (1993). From both definitions \( u_t \) is linear innovation of \( \epsilon_t^2 \), however, in the Drost and Nijman approach, \( u_t \) is orthogonal to all past values of \( \epsilon_t \). This additional constraint ensures the stability of the class under temporal aggregation, however our model uses contemporaneous aggregation, not temporal aggregation, thus we don’t need this extra assumption.

**Theorem 40.** (Contemporaneous aggregation of GARCH processes) The sum of two fourth order stationary GARCH\((1,1)\) processes with independent disturbances is weak GARCH\((2,2)\).

**Proof.** Consider the sum of two GARCH\((1,1)\) processes \( (\epsilon_{1,t}) \) and \( (\epsilon_{2,t}) \) i.e.

\[
(\epsilon_t) \text{ satisfies that } \epsilon_t = \epsilon_{1,t} + \epsilon_{2,t}, \text{ for all } t \in \mathbb{Z}.
\]

where

\[
\epsilon_{1,t} = \sigma_{1,t} \eta_{1,t}, \quad \sigma_{1,t}^2 = \omega + \alpha_i \epsilon_{1,t-1}^2 + \beta_i \sigma_{1,t-1}^2, \quad \omega_1 > 0, \alpha_i, \beta_i \geq 0, \quad (\eta_{1,t}) \text{ i.i.d. } (0,1), \quad i = 1, 2.
\]

and suppose that the sequences \( (\eta_{1,t}) \) and \( (\eta_{2,t}) \) are independent.

Note that \( (\epsilon_{1,t}) \) and \( (\epsilon_{2,t}) \) are centered, uncorrelated and mutually independent processes, due to the independence assumptions imposed on \( (\eta_{1,t}) \) and \( (\eta_{2,t}) \), therefore \( \mathbb{E}(\epsilon_t) = 0 \) and \( \text{Cov}(\epsilon_t, \epsilon_{t-h}) = 0 \) for all \( h > 0, t \in \mathbb{Z} \), therefore \( (\epsilon_t) \) is white noise. Moreover, for \( h > 0 \) we have that:

\[
\text{Cov}(\epsilon_{1,t}^2, \epsilon_{2,t-h}^2) = 0,
\]

\[
\text{Cov}(\epsilon_{1,t} \epsilon_{2,t}, \epsilon_{1,t-1}^2) = \mathbb{E}(\epsilon_{1,t} \epsilon_{2,t} \epsilon_{1,t-1}^2) = \mathbb{E}(\epsilon_{1,t}) \mathbb{E}(\epsilon_{2,t} \epsilon_{1,t-1}^2) = 0.
\]

Consequently, for \( h > 0 \)

\[
\text{Cov}(\epsilon_t^2, \epsilon_{t-h}^2) = \text{Cov}(\epsilon_{1,t}^2, \epsilon_{1,t-h}^2) + \text{Cov}(\epsilon_{2,t}^2, \epsilon_{2,t-h}^2).
\]
If \((\epsilon_{i,t})\) is fourth order stationary i.e. \(E(\epsilon_{i,t}^4) = \kappa_i < \infty\) then \((\epsilon_{i,t}^2)\) follows an ARMA(1,1) process. In particular, the autocovariances of the squares take the form
\[
\gamma_{i,t}^2(h) = Cov(\epsilon_{i,t}^2, \epsilon_{i,t-h}^2) = \gamma_{i,t}^2(1)(\alpha_i + \beta_i)h^{-1}, \quad h \geq 1. 
\tag{3.44}
\]
Combining equations (3.43) and (3.44) we obtain that
\[
\gamma_{i,t}^2(h) = Cov(\epsilon_{i,t}^2, \epsilon_{i,t-h}^2) = \gamma_{i,t}^2(1)(\alpha_1 + \beta_1)h^{-1} + \gamma_{i,t}^2(1)(\alpha_2 + \beta_2)h^{-1}. 
\tag{3.45}
\]
Given a function \(f\) defined on the integers, \(Lf(h) = f(h-1), h > 0\). Therefore, \((1 - \beta L)^h = 0\) for \(h > 0\). Hence,
\[
[1 - (\alpha_1 + \beta_1 L)][1 - (\alpha_2 + \beta_2 L)]\gamma_{i,t}(h) = 0 \quad h > 2.
\]
From Theorem 36, we obtain that \((\epsilon_i^2)\) is a weak GARCH(2,2) process of the form
\[
[1 - (\alpha_1 + \beta_1 L)][1 - (\alpha_2 + \beta_2 L)]\epsilon_i^2 = \omega + u_i + \theta_1 u_{i-1} + \theta_2 u_{i-2},
\]
where \((u_i)\) is the linear innovation process of \((\epsilon_i^2)\). The orders obtained in the ARMA representation of \((\epsilon_i^2)\) are not necessarily the minimum ones. Note that if \(\alpha_1 + \beta_1 = \alpha_2 + \beta_2\) then \(\gamma_{i,t}(h) = \gamma_{i,t}^2(1)(\alpha_1 + \beta_1)h^{-1}\) for \(h \geq 1\). Therefore, \([1 - (\alpha_1 + \beta_1 L)]\gamma_{i,t}(h) = 0\) if \(h > 1\). Hence, \((\epsilon_i^2)\) is a weak GARCH(1,1) process.

Remark. This proof can be found in Franq, C. and Zakoian, J. (2010).

The class of weak ARMA-GARCH models is an extension of the class of weak GARCH models. The following definition was provided by Franq, C. and Zakoian, J. (2000).

**Definition 3.5.2 (Weak ARMA-GARCH).** A process \((X_t)\) is said to be **weak ARMA-GARCH** if it satisfies the following two stage representation:
\[
X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \epsilon_t + \sum_{j=1}^{Q} \psi_j \epsilon_{t-j}, \\
\epsilon_t^2 - \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 = \omega + u_t + \sum_{j=1}^{q} \beta_j u_{t-j},
\tag{3.46}
\]
where the two polynomials \(\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p\) and \(\Psi(z) = 1 + \psi_1 z + \cdots + \psi_Q z^Q\) have all their zeros outside the unit disk and have no common zeros, the same regularity assumptions are made about the polynomials \(\phi(z) = 1 - \alpha_1 z - \cdots - \alpha_p z^p\) and \(\psi(z) = \omega + \beta_1 z + \cdots + \beta_q z^q\), and \((\epsilon_t)\) and \((u_t)\) are linear innovation processes of \((X_t)\) and \((\epsilon_t^2)\) respectively.

The following procedure was proposed by Francq and Zakoian to estimate the parameters of (3.46). First, define \(\theta^{(1)} = (\phi_1, \cdots, \phi_p, \psi_1, \cdots, \psi_Q)\), \(\theta^{(2)} = (\omega, \alpha_1, \cdots, \alpha_p, \beta_1, \cdots, \beta_q)\) and \(\theta = (\theta^{(1)}, \theta^{(2)})\). Let \(X_1, \cdots, X_n\) be a realization of length \(n\) of \((X_t)\). For \(0 < t < n\), \(\epsilon_t(\theta)\) and \(u_t(\theta)\) are approximately by \(\tilde{\epsilon}(\theta)\) and \(\tilde{u}(\theta)\) obtained by replacing the unknown starting values by zero \((\epsilon_1(\theta) = 0, -Q + 1 \leq t < 0\) and \(u_1(\theta) = 0, -q + 1 \leq t < 0\) ). A least squares estimator \(\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)\) can be obtained using the following procedure: Minimize \(Q_n^{(1)}(\theta_1)\), where \(Q_n^{(1)}(\theta_1) = \frac{1}{n} \sum_{t=1}^{n} \epsilon_t^2(\theta_1)\) and let \(\hat{\theta}_1 = \arg \min_{\theta_1} Q_n^{(1)}(\theta_1)\), then, minimize \(Q_n^{(2)}(\theta_2)\) of \(\theta_2\) and let \(\hat{\theta}_2 = \arg \min_{\theta_2} Q_n^{(2)}(\theta_2)\). Franq, C. and Zakoian, J. (2000) showed that under certain regularity assumptions the estimator \(\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)\) is consistent.

It is time now to introduce our last model, which assumes that both the series of centered excess market returns
(\epsilon_{m,t}) and the series of centered excess asset returns (\epsilon_{f,t}) follow fourth order stationary and mutually independent GARCH(1,1) processes.

For each \( s \geq \tau + 1 \) we have that:

\[
\begin{align*}
R_{f,t} &= \alpha_s + \beta_s R_{m,t} + u_t, \text{ for all } t = s - \tau, \ldots, s - 1. \\
\epsilon_{f,t} &= \sqrt{h_{f,t}} z_{f,t}, \\
h_{f,t} &= \omega_{f,s} + \alpha_{1,f,s} \epsilon_{f,t-1}^2 + \beta_{1,f,s} h_{f,t-1}, \\
\epsilon_{m,t} &= \sqrt{h_{m,t}} z_{m,t}, \\
h_{m,t} &= \omega_{m,s} + \alpha_{1,m,s} \epsilon_{m,t-1}^2 + \beta_{1,m,s} h_{m,t-1}.
\end{align*}
\]

(3.47)

where \( R_{m,t} = \epsilon_{m,t} + \mu_{m,s} \) and \( R_{f,t} = \epsilon_{f,t} + \mu_{f,s} \) (note that \( \mu_{m,s} = \mathbb{E}(R_{m,t}) \) and \( \mu_{f,s} = \mathbb{E}(R_{f,t}) \)), \((z_{m,t})\) and \((z_{f,t})\) are i.i.d. and mutually independent with \( \mathbb{E}(z_{m,t}) = 0, \text{Var}(z_{m,t}) = 1 \) and \( \mathbb{E}(z_{f,t}) = 0, \text{Var}(z_{f,t}) = 1 \) respectively. \((u_t)\) is assumed to be centered i.e. \( \mathbb{E}(u_t) = 0 \), also \( \mathbb{E}(\epsilon_{m,1}^4) = \kappa_m < \infty \) and \( \mathbb{E}(\epsilon_{f,1}^4) = \kappa_f < \infty \), for all \( t = s - \tau, \ldots, s - 1 \).

From (3.47), we have that

\[
\epsilon_{f,t} + \mu_{f,s} = \alpha_s + \beta_s (\epsilon_{m,t} + \mu_{m,s}) + u_t.
\]

Therefore,

\[
\begin{align*}
u_t &= (\epsilon_{f,t} - \beta_s \epsilon_{m,t}) + (\mu_{f,s} - (\alpha_s + \beta_s \mu_{m,s})) \\
&= (\epsilon_{f,t} - \beta_s \epsilon_{m,t}),
\end{align*}
\]

(3.48)

because \((u_t)\) is assumed to be centered (this is a standard assumption of factor models). From Theorem 40 we have that the process \( \epsilon_{f,t} - \beta_s \epsilon_{m,t} \) is weak GARCH(2,2). Therefore the residuals \((u_t)\) can be written as a weak GARCH(2,2) process.

To estimate the parameters \( \alpha_s \) and \( \beta_s \) we follow a procedure similar to the one proposed by Francq and Zakoian to estimate the parameters of a weak ARMA-GARCH model. First, the sample mean of both series of returns are computed: \( \hat{\mu}_{m,s} = \frac{1}{T} \sum_{t=s-\tau}^{s-1} R_{m,t} \) and \( \hat{\mu}_{f,s} = \frac{1}{T} \sum_{t=s-\tau}^{s-1} R_{f,t} \). Let \( \epsilon_{m,t} = R_{m,t} - \hat{\mu}_{m,s} \) and \( \epsilon_{f,t} = R_{f,t} - \hat{\mu}_{f,s} \) and define \( u_t = \epsilon_{f,t} - \beta_s \epsilon_{m,t}, \) for \( s - \tau \leq t \leq s - 1 \), then there exist real coefficients \( \omega_s, \alpha_{1,s}, \alpha_{2,s}, \beta_{1,s}, \beta_{2,s} \) such that

\[
u_t^2 = \alpha_{1,s} \nu_{t-1}^2 + \alpha_{2,s} \nu_{t-2}^2 + \omega_s + \eta_t + \beta_{1,s} \eta_{t-1} + \beta_{2,s} \eta_{t-2}, \quad \text{with } \eta_t \text{ linear innovation of } \nu_t^2.
\]

Let \( \theta_s = (\omega_s, \alpha_{1,s}, \alpha_{2,s}, \beta_{1,s}, \beta_{2,s}) \), we approximate \( u_t(\beta_s) \) and \( \eta_t(\theta_s) \) by \( \tilde{u}_t(\beta_s) \) and \( \tilde{\eta}_t(\theta_s) \) where \( \tilde{u}_t(\beta) = 0 \) and \( \tilde{\eta}_t(\theta_s) = 0 \) for \( t = s - (\tau + 2), s - (\tau + 1) \). Our least squares estimator of \( \phi_s = (\beta_s, \theta_s) \) is obtained by minimizing \( Q_\tau(\phi_s) = \frac{1}{T} \sum_{t=s-\tau}^{s-1} \tilde{\eta}_t^2(\phi_s) \) i.e. \( \tilde{\phi}_s = \arg \min_{\phi_s \in \Phi_s} Q_\tau(\phi_s) \).

Note that

\[
\begin{align*}
\tilde{\eta}_{s-\tau}(\phi_s) &= \tilde{u}_{s-\tau}^2(\phi_s) - \omega_s - (\epsilon_{f,s-\tau} - \beta_s \epsilon_{m,s-\tau})^2 - \omega_s \\
\tilde{\eta}_{s-(\tau-1)}(\phi_s) &= (\epsilon_{f,s-(\tau-1)} - \beta_s \epsilon_{m,s-(\tau-1)})^2 - (\alpha_{1,s} + \beta_{1,s}) (\epsilon_{f,s-\tau} - \beta_s \epsilon_{m,s-\tau})^2 + \beta_{1,s} \omega_s - \omega_s \\
\eta_t &= u_t^2 - (\alpha_{1,s} \nu_{t-1}^2 + \alpha_{2,s} \nu_{t-2}^2 + \omega_s + \beta_{1,s} \eta_{t-1} + \beta_{2,s} \eta_{t-2}), \quad \text{for all } s - (\tau - 2) \leq t \leq s - 1.
\end{align*}
\]

From our assumption of residuals \((u_t)\) being centered we obtain that \( \mu_{f,s} - (\alpha_s + \beta_s \mu_{m,s}) = 0 \). Thus \( \hat{\alpha}_s = \hat{\mu}_{f,s} - \beta_s \hat{\mu}_{m,s} \), where \( \beta_s \) was obtained in the previous step.
3.6 Applications to hedge fund return modelling

A substantial growth in the number of hedge funds has occurred since the 1990s, they have become popular for their ability to generate an absolute return at reduced risk regardless of the market conditions. The class of hedge fund strategies is very heterogeneous. Hedge funds can be classified into a number of different strategy groups depending on the main strategy followed. According to Fuss, R., Kaiser, D. and Adams, Z. (2007) the three main categories are: arbitrage, event driven, and directional.

Arbitrage strategies try to take advantage of temporarily wrong valuations between different financial instruments and attempt to offer their investors very low market exposure. Event driven strategies try to take advantage of price anomalies triggered by pending or upcoming firm transactions such as mergers, restructurings, liquidations and insolvencies. The success of these strategies comes from a false judgment of a situation and the uncertainty of other investors in the case of takeovers, reorganizations, or management buyouts. Directional strategies are characterized by significant market exposure (long and short). These three groups can be subdivided into more specific strategies. See Fuss, R., Kaiser, D. and Adams, Z. (2007).

The merger arbitrage strategy belongs to the event driven management style class, it invests simultaneously in long and short positions by purchasing the stocks of the company being taken over and selling those of the take-over company. Generally, stocks that are object of a take-over gain in value while stocks of the take-over company fall in value. An interesting fact mentioned in the literature is that the merger arbitrage strategy displays high correlation to the equity markets when the latter declines and, in turn, displays low correlation when stocks trade up or sideways. See Mitchell, M. and Pulvino, T. (2001).

The equity hedge or long/short equity strategy belongs to the directional management style class. It involves buying long equities that are expected to increase in value and selling short equities that are expected to decrease in value. Many equity hedge funds have a long bias. Besides stocks, equity hedge funds can also be invested in other assets on a limited scale. See Kat, H. M. and Lu, S. (2002), Fung, W. and Hsieh, D. (2004) and Fuss, R., Kaiser, D. and Adams, Z. (2007).

In this section we study the market exposure and performance of three different hedge fund strategies: equity hedge, event driven, and merger arbitrage, using the time-varying factor models described in the previous section. According to Wei, W. (2010) it is generally accepted that hedge fund returns can be separated into alpha and beta components. The first one attributable to manager skill and the second one being generated by market exposure. Modelling hedge fund returns involves primarily estimating the magnitude and direction of these alpha and beta components.

The considered 1-factor models with time-varying parameters use a rolling window regression to account for dynamic changes in the model’s parameters. Time-varying alpha and beta at time $t$ are determined by regressing return series from $t - \tau$ to $t - 1$. Each of the models has the general form:

For each $t \geq \tau + 1$,

$$R_{f,t-j} = \alpha_t + \beta_t R_{m,t-j} + \epsilon_{t-j}, \quad \text{for all } j = 1, \cdots, \tau,$$

(3.49)

where $\tau$ is the window length, $R_{f,t-j} = r_{f,t-j} - r_{F,t-j}$ and $R_{m,t-j} = r_{m,t-j} - r_{F,t-j}$ where $r_{f,t-j}$, $r_{m,t-j}$ and $r_{F,t-j}$ denote the daily return of a hedge fund strategy, the daily return of the market and the daily risk free rate respectively, at time $t - j$.

Model 1 assumes that residuals $(\epsilon_{t-j})$ are i.i.d. such that $\epsilon_{t-j} \sim N(0, \sigma^2_{\epsilon,t})$ for all $j = t - \tau, \cdots, t - 1$. This is a standard assumption of macroeconomic factor models, however, we consider different assumptions in the remaining models. Model 2 assumes that the series of residuals $(\epsilon_{t-j})$ is a GARCH(1,1) process. Model 3
assumes that the series of centered excess market returns \( (R_{m,t-j} - \mu_{m,t}) \) is a GARCH(1,1) process and the series of residuals \( (\epsilon_{t-j}) \) is i.i.d. such that \( \epsilon_{t-j} \sim N(0, \sigma_{\epsilon,t}^2) \) for all \( j = t - \tau, \ldots, t - 1 \). Model 4 assumes that the series of residuals \( (\epsilon_{t-j}) \) is a weak GARCH(2,2) process. Time-varying alpha and beta of three hedge fund strategies are computed using each of these models, and results are compared.

As proxies for the three hedge fund strategies considered (equity hedge or long/short, event driven and merger arbitrage), the daily prices of the following indices from HFR (http://www.hedgefundresearch.com) are used: HFRX Equity Hedge Index, HFRX Event Driven Index and HFRX ED: Merger Arbitrage Index. The Standard & Poor’s 500 Index (S&P 500) is used as a proxy for the market returns, the daily prices of the S&P 500 were obtained from Yahoo Finance (https://finance.yahoo.com/quote/%5EGSPC/history). Our sample contains 3377 observations of daily prices of each hedge fund index and of the S&P 500 Index, the first observation corresponds to March 31st, 2003 and the last one to September 5th, 2017. In all our models the explanatory and the response variables are excess returns. The excess return of an asset at time \( t \) is defined as \( R_t = r_t - r_{F,t} \), where \( r_t \) and \( r_{F,t} \) denote the return of the asset and the risk free rate at time \( t \) respectively. The 3 month T-bill rate was used as a proxy for the risk free rate (https://fred.stlouisfed.org/series/DTB3). We used a 253 days (approximately 1 year of business days) rolling window.

Figure 3.1 shows the graphs of time-varying betas of the equity hedge strategy for three of the four models previously discussed in section 3.5: models 1, 2 and 4. Note that models 2.1 and 2.2 both refer to model 2, the difference lies in the estimation method. Model 2.1 uses a three step least squares procedure while the parameters of model 2.2 are estimated using quasi-loglikelihood. Model 1 assumes that the series of residuals is a Gaussian white noise process while model 2 assumes GARCH(1,1) residuals. Figure 3.1 shows that the graphs of models 1, 2.1 and 2.2 look quite similar and do not seem to capture
Figure 3.2: Time-varying alpha of the equity hedge strategy for models 1, 2 and 4.

recent exposures, which suggests that considering GARCH residuals versus Gaussian white noise residuals does not change significantly time-varying beta estimates. Note that we did not include the graph of model 3 in figure 3.1 because the time-varying betas obtained using this model, move in a “weird” way that does not reflect the behavior of hedge fund exposures to markets. The time-varying betas of each strategy, obtained using model 3, are shown in figure A.5. On the other hand, model 4 is able to capture recent market exposures and to react faster to market changes, this becomes evident during the subprime mortgage crisis (2008 crisis).

To illustrate our point, we quote the article *Global financial crisis: five key stages 2007-2011* authored by Elliot, L. (2011, August 7) and published by the British newspaper *The Guardian*:

“... 9 August 2007. 15 September 2008. 2 April 2009. 9 May 2010. 5 August 2011. From sub-prime to down-grade, the five stages of the most serious crisis to hit the global economy since the Great Depression can be found in those dates.

Phase one on 9 August 2007 began with the seizure in the banking system precipitated by BNP Paribas announcing that it was ceasing activity in three hedge funds that specialised in US mortgage debt. This was the moment it became clear that there were tens of trillions of dollars worth of dodgy derivatives swilling round which were worth a lot less than the bankers had previously imagined. Nobody knew how big the losses were or how great the exposure of individual banks actually was, so trust evaporated overnight and banks stopped doing business with each other.

It took a year for the financial crisis to come to a head but it did so on 15 September 2008 when the US government allowed the investment bank Lehman Brothers to go bankrupt. Up to that point, it had been assumed that governments would always step in to bail out any bank that got into serious trouble...

When Lehman Brothers went down, the notion that all banks were “too big to fail” no longer held true, with the
result that every bank was deemed to be risky..."

The only graph that is able to capture the sudden exposure drop that happened as a consequence of these two events from the 2008 crisis, corresponds to model 4 (which assumes that residuals are weak GARCH(2,2)), which suggests that this model unlike the other three is able to capture recent market exposures. Similar results were observed for the other 2 hedge fund strategies. See figures A.1 and A.3.

Figure 3.2 shows the time-varying alphas of the equity hedge strategy, obtained using models 1, 2 and 4. The graphs look quite similar except during the 2008 crisis period, where model 2 shows a higher alpha than models 1 and 4. The time-varying alphas of each strategy for model 3 are shown in figure A.6.

Our three hedge fund strategies were compared using the graphs of the time-varying betas and alphas produced by model 4 (figures 3.3 and 3.4). The average alpha and beta over three different sub-periods were computed for each strategy: March 31st, 2004 - July 31st, 2007 (post dotcom bubble), August 1st, 2007 - June 30th, 2009 (subprime mortgage crisis), July 1st, 2009 - September 6th, 2017 (post financial crisis). See tables 3.1 and 3.2.

Figure 3.3 reveals some interesting trends. Note that while the largest exposures of the equity hedge strategy and the event driven strategy occurred during the post dotcom bubble, the highest exposure of the merger arbitrage strategy happened during the 2008 crisis. Overall, equity hedge exhibits the highest market exposure among all the strategies, followed by event driven and merger arbitrage, in that order. The results from table 3.2 confirm it. Also, during the subprime mortgage crisis, the market exposures of the different strategies where, on average, closer to each other than during the post dotcom bubble and after the 2008 financial crisis, see table 3.2. In addition, from November 2008 to October 2009 the merger arbitrage strategy exhibits a higher market exposure than the rest; this is not surprising as this type of hedge fund takes advantage of mergers and acquisitions.

Figure 3.4 shows the graphs of time-varying alphas for each strategy. During the year 2008 the merger arbitrage alpha shows an increasing trend, while the opposite happens to the other two strategies. Monday August 8, 2011, marked the day the US and global stock markets crashed following the Friday night credit rating downgrade by Standard and Poor’s of the United States sovereign debt from AAA, or “risk free”, to AA+. It was the first time in history the United States was downgraded. Note that a decline of the equity hedge alpha started at the beginning of August of 2011 and lasted until mid January of 2012.

Table 3.2 shows that the average equity hedge alpha was negative every sub-period. The event driven average alpha was negative only during the subprime mortgage crisis and positive during the other two sub-periods, while the merger arbitrage average alpha was positive in every sub-period. These results reveal that the event driven strategy, and more so the merger arbitrage strategy outperform the equity hedge strategy in the generation of positive alpha, specially in turbulent and adverse markets.

### 3.7 Conclusion

The assumptions and estimation methods, of four different 1-factor models with time-varying parameters have been discussed in this chapter. Using data of three different hedge fund strategies and a market index, the time-
Table 3.2: Average beta per period

<table>
<thead>
<tr>
<th>Event</th>
<th>Beta</th>
<th>Equity hedge</th>
<th>Event driven</th>
<th>Merger arbitrage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post dotcom bubble</td>
<td>0.408%</td>
<td>0.26076%</td>
<td>0.14129%</td>
<td></td>
</tr>
<tr>
<td>Subprime mortgage crisis</td>
<td>0.2679%</td>
<td>0.1954%</td>
<td>0.24134%</td>
<td></td>
</tr>
<tr>
<td>Post financial crisis</td>
<td>0.3072%</td>
<td>0.211467%</td>
<td>0.0981%</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.3: Time-varying beta from model 4 for each strategy.

Figure 3.4: Time-varying alpha from model 4 for each strategy.
varying alpha and beta of each model have been estimated. The first model assumes that the series of residuals is a Gaussian white noise process, while the second one assumes that the series of residuals is a GARCH(1,1) process, the third one uses an unobserved GARCH process to model the asset excess return and the last one assumes that the series of residuals is a weak GARCH(2,2) process. The model that performed the best was the one with weak GARCH(2,2) residuals, because unlike the other three, it was able to capture recent market exposures. To the best of our knowledge, this model has not been used before to estimate the time-varying alpha and beta of factor models.
Appendix A

Graphs

Figure A.1: Time-varying beta of the event driven strategy for models 1, 2 and 4.
Figure A.2: Time-varying alpha of the event driven strategy for models 1, 2 and 4.

Figure A.3: Time-varying beta of the merger arbitrage strategy for models 1, 2 and 4.
Figure A.4: Time-varying alpha of the merger arbitrage strategy for models 1, 2 and 4.

Figure A.5: Time-varying beta of each strategy for model 3.
Figure A.6: Time-varying alpha of each strategy for model 3.
Bibliography


